



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 9/10, 9/90, 9/92, 15/54, 15/61	A1	(11) International Publication Number: WO 98/50531
		(43) International Publication Date: 12 November 1998 (12.11.98)

(21) International Application Number: PCT/AU98/00315	(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).
(22) International Filing Date: 1 May 1998 (01.05.98)	
(30) Priority Data: PO 6545 1 May 1997 (01.05.97) AU PO 8162 22 July 1997 (22.07.97) AU	
(71) Applicant (for all designated States except US): THE UNIVERSITY OF SYDNEY [AU/AU]; Parramatta Road, Sydney, NSW 2006 (AU).	
(72) Inventors; and	Published
(75) Inventors/Applicants (for US only): REEVES, Peter, Richard [GB/AU]; 20 Mansfield Street, Glebe, NSW 2037 (AU). WANG, Lei [AU/AU]; 8A Holt Street, North Ryde, NSW 2113 (AU).	With international search report.
(74) Agent: GRIFFITH HACK; G.P.O. Box 4164, Sydney, NSW 2001 (AU).	

(54) Title: NUCLEIC ACID MOLECULES SPECIFIC FOR BACTERIAL ANTIGENS AND USES THEREOF

(57) Abstract

The present invention relates to nucleic acid molecules derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit, including a *wzx* gene or a *wzy* gene, or a gene with a similar function; the gene being involved in the synthesis of a particular bacterial polysaccharide antigen, wherein the sequence of the nucleic acid molecule is specific to the particular bacterial polysaccharide antigen. Polysaccharides to which the invention relates include O antigens. The invention also relates to methods of testing samples for the presence of one or more bacterial polysaccharide antigens, using the nucleic acid molecules of the invention, and to kits containing the nucleic acid molecules of the invention.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		

- 1 -

Nucleic acid molecules specific for bacterial
antigens and uses thereof.

TECHNICAL FIELD

5 The invention relates to novel nucleotide sequences located in a gene cluster which controls the synthesis of a bacterial polysaccharide antigen, especially an O antigen, and the use of those nucleotide sequences for the detection of bacteria which express particular 10 polysaccharide antigens (particularly O antigens) and for the identification of the polysaccharide antigens (particularly O antigens) of those bacteria.

BACKGROUND ART

15 Enteropathogenic E. coli strains are well known causes of diarrhoea and haemorrhagic colitis in humans and can lead to potentially life threatening sequelae including haemolytic uremic syndrome and thrombotic thrombocytopenic purpura. Some of these strains are 20 commonly found in livestock and infection in humans is usually a consequence of consumption of contaminated meat or dairy products which have been improperly processed. The O specific polysaccharide component (the "O antigen") of lipopolysaccharide is known to be a major virulence 25 factor of enteropathogenic E. coli strains.

The E. coli O antigen is highly polymorphic and 166 different forms of the antigen have been defined; Ewing, W. H. [in Edwards and Ewings "Identification of the Enterobacteriaceae" Elsevier. Amsterdam (1986)] discusses 30 128 different O antigens while Lior H. (1994) extends the number to 166 [in "Classification of *Escherichia coli* In *Escherichia coli* in domestic animals and humans pp31-72. Edited by C.L.Gyles CAB International]. The genus Salmonella enterica has 46 known O antigen types [Popoff 35 M.Y. et al (1992) "Antigenic formulas of the Salmonella enterica serovars" 6th revision WHO Collaborating Centre for Reference and Research on Salmonella enterica, Institut Pasteur Paris France].

- 2 -

An important step in determining the biosynthesis of O antigens and therefore the mechanism of the polymorphism has been to characterise the gene clusters controlling O antigen biosynthesis. The genes specific for the synthesis of the O antigen are generally located in a gene cluster at map position 45 minutes on the chromosome of E. coli K-12 [Bachmann, B. J. 1990 "Linkage map of Escherichia coli K-12". *Microbiol. Rev.* 54: 130-197], and at the corresponding position in S. enterica LT2 [Sanderson et al (1995) "Genetic map of Salmonella enterica typhimurium", Edition VIII *Microbiol. Rev.* 59: 241-303]. In both cases the O antigen gene cluster is close to the *gnd* gene as is the case in other strains of E. coli and S. enterica [Reeves P.R. (1994) "Biosynthesis and assembly of lipopolysaccharide", 281-314. in A. Neuberger and L.L.M. van Deenen (eds) "Bacterial cell wall, new comprehensive biochemistry" vol 27 Elsevier Science Publishers]. These genes encode enzymes for the synthesis of nucleotide diphosphate sugars and for assembly of the sugars into oligosaccharide units and in general for polymerisation to O antigen.

The E. coli O antigen gene clusters for a wide range of E. coli O antigens have been cloned but the O7, O9, O16 and O111 O antigens have been studied in more detail with only O9 and O16 having been fully characterised with regard to nucleotide sequence to date [Kido N., Torgov V.I., Sugiyama T., Uchiya K., Sugihara H., Komatsu T., Kato N. & Jann K. (1995) "Expression of the O9 polysaccharide of Escherichia coli: sequencing of the E. coli O9 rfb gene cluster, characterisation of mannosyl transferases, and evidence for an ATP-binding cassette transport system" *J. of Bacteriol.* 177 2178-2187; Stevenson G., Neal B., Liu D., Hobbs M., Packer N.H., Batley M., Redmond J.W., Lindquist L. & Reeves PR (1994) "Structure of the O antigen of E. coli K12 and the sequence of its rfb gene cluster" *J. of Bacteriol.* 176 4144-4156; Jayaratne, P. et al. (1991) "Cloning and analysis of duplicated rfbM and rfbK genes involved in the

formation of GDP-mannose in *Escherichia coli* O9:K30 and participation of *rfb* genes in the synthesis of the group 1 K30 capsular polysaccharide" *J. Bacteriol.* 176: 3126-3139; Valvano, M. A. and Crosa, J. H. (1989) "Molecular cloning and expression in *Escherichia coli* K-12 of chromosomal genes determining the O7 lipopolysaccharide antigen of a human invasive strain of *E. coli* O7:K1". *Inf and Immun.* 57:937-943; Marolda C. L. And Valvano, M. A. (1993). "Identification, expression, and DNA sequence of the GDP-mannose biosynthesis genes encoded by the O7 *rfb* gene cluster of strain VW187 (*Escherichia coli* O7:K1)". *J. Bacteriol.* 175:148-158.]

Bastin D.A., et al. 1991 ["Molecular cloning and expression in *Escherichia coli* K-12 of the *rfb* gene cluster determining the O antigen of an *E. coli* O111 strain". *Mol. Microbiol.* 5:9 2223-2231] and Bastin D.A. and Reeves, P.R. [(1995) "Sequence and analysis of the O antigen gene(*rfb*)cluster of *Escherichia coli* O111". *Gene* 164: 17-23] isolated chromosomal DNA encoding the *E. coli* 0111 *rfb* region and characterised a 6962 bp fragment of *E. coli* 0111 *rfb*. Six open reading frames (orfs) were identified in the 6962 bp partial fragment and the alignment of the sequences of these orfs revealed homology with genes of the GDP-mannose pathway, *rfbK* and *rfbM*, and other *rfb* and *cps* genes.

The nucleotide sequences of the loci which control expression of *Salmonella enterica* B, A, D1, D2, D3, C1, C2 and E O antigens have been characterised [Brown, P. K., L. K. Romana and P. R. Reeves (1991) "Cloning of the *rfb* gene cluster of a group C2 *Salmonella enterica*: comparison with the *rfb* regions of groups B and D *Mol. Microbiol.* 5:1873-1881; Jiang, X.-M., B. Neal, F. Santiago, S. J. Lee, L. K. Romana, and P. R. Reeves (1991) "Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella enterica* serovar *typhimurium* (LT2)". *Mol. Microbiol.* 5:692-713; Lee, S. J., L. K. Romana, and P. R. Reeves (1992) "Sequences and structural analysis of the *rfb* (O antigen)gene cluster from a group C1 *Salmonella enterica*

- 4 -

enterica strain" J. Gen. Microbiol. **138**: 1843-1855; Lui, D., N. K. Verma, L. K. Romana, and P. R. Reeves (1991) "Relationship among the *rfb* regions of Salmonella enterica serovars A, B and D" J. Bacteriol. **173**: 4814-4819; Verma, 5 N. K., and P. Reeves (1989) "Identification and sequence of *rfbS* and *rfbE*, which determine the antigenic specificity of group A and group D Salmonella entericae" J. Bacteriol. **171**: 5694-5701; Wang, L., L. K. Romana, and P. R. Reeves (1992) "Molecular analysis of a Salmonella 10 enterica enterica group E1 *rfb* gene cluster: O antigen and the genetic basis of the major polymorphism" Genetics **130**: 429-443; Wyk, P., and P. Reeves (1989). "Identification and sequence of the gene for abequose synthase, which confers antigenic specificity on group B 15 Salmonella entericae: homology with galactose epimerase" J. Bacteriol. **171**: 5687-5693; Xiang, S. H., M. Hobbs, and P. R. Reeves. 1994 Molecular analysis of the *rfb* gene luster of a group D2 Salmonella enterica strain: evidence for its origin from an insertion sequence -mediated 20 recombination event between group E and D1 strains. J. Bacteriol. **176**: 4357 -4365; Curd, H., D. Liu and P. R. Reeves, 1998. Relationships among the O antigen Salmonella enterica groups B, D1, D2, and D3. J. Bacteriol. **180**: 1002-1007.].

25 Of the closely related Shigella (which really can be considered to be part of E. coli) S. dysenteriae and S. flexneri O antigens have been fully sequenced and are next to *gnd*. [Klena JD & Schnaitman CA (1993) "Function of the *rfb* gene cluster and the *rfe* gene in the synthesis of O 30 antigen by *Shigella dysenteriae* 1" Mol. Microbiol. **9** 393-402; Morona R., Mavris M., Fallarino A. & Manning P. (1994) "Characterisation of the *rfc* region of *Shigella flexneri*" J.Bacteriol **176**: 733-747]

35 Inasmuch as the O antigen of enteropathogenic E. coli strains and the O antigen of Salmonella enterica strains are major virulence factors and are highly polymorphic, there is a real need to develop highly specific, sensitive, rapid and inexpensive diagnostic assays to

detect E. coli and assays to detect S. enterica. There is also a real need to develop diagnostic assays to identify the O antigens of E. coli strains and assays to identify the O antigens of S. enterica strains. With regard to the 5 detection of E. coli these needs extend beyond EHEC (enteropathogenic haemorrhagic E. coli) strains but this is the area of greatest need. There is interest in diagnostics for ETEC (enterotoxigenic E. coli) etc in E. coli.

10 The first diagnostic systems employed in this field used large panels of antisera raised against E. coli O antigen expressing strains or S. enterica O antigen expressing strains. This technology has inherent difficulties associated with the preparation, storage and 15 usage of the reagents, as well as the time required to achieve a meaningful diagnostic result.

Nucleotide sequences derived from the O antigen gene clusters of S. enterica strains have been used to determine S. enterica O antigens in a PCR assay [Luk, 20 J.M.C. et al. (1993) "Selective amplification of abequose and paratose synthase genes (rfb) by polymerase chain reaction for identification of S. enterica major serogroups (A, B, C2, and D)", *J. Clin. Microbiol.* 31:2118-2123]. The prior complete nucleotide sequence characterisation of 25 the entire rfb locus of serovars Typhimurium, Paratyphi A, Typhi, Muenchen, and Anatum; representing groups B, A, D1, C2 and E1 respectively enabled Luk et al. to select oligonucleotide primers specific for those serogroups. Thus the approach of Luk et al. was based on aligning 30 known nucleotide sequences corresponding to CDP-abequose and CDP-paratose synthesis genes within the O antigen regions of S. enterica serogroups E1, D1, A, B and C2 and exploiting the observed nucleotide sequence differences in order to identify serotype-specific oligonucleotides.

35 In an attempt to determine the O antigen serotype of a Shiga-like toxin producing E. coli strain, Paton, A. W., et al. 1996 ["Molecular microbiological investigation of an outbreak of Hemolytic-Uremic Syndrome caused by dry

fermented sausage contaminated with Shiga-like toxin producing *Escherichia coli*". *J. Clin. Microbiol.* 34: 1622-1627], used oligonucleotides derived from the *wbdI* (orf6) region, which were believed to be specific to the *E. coli* 5 0111 antigen and which were derived from *E. coli* 0111 sequence, in a PCR diagnostic assay. Unpublished reports indicate that the approach of Paton et al. is deficient in that the nucleotide sequences derived from *wbdI* may not 10 specifically identify the 0111 antigen and in fact lead to detection of false positive results. Paton et al. disclose the detection of 5 0111 antigen isolates by PCR when in fact from only 3 of those isolates did they detect bacteria which reacted with 0111 specific antiserum.

15 **DESCRIPTION OF THE INVENTION**

Whilst not wanting to be held to a particular hypothesis, the present inventors now believe that the reported false positives found with the Paton et al. 20 method are due to the fact that the nucleic acid molecules employed by Paton et al. were derived from genes which have a putative function as a sugar pathway gene, [Bastin D.A. and Reeves, P.R. (1995) Sequence and analysis of the O antigen gene(*rfb*) cluster of *Escherichia coli* 0111. *Gene* 164: 17-23] which they now believe to lack the necessary 25 nucleotide sequence specificity to identify the *E. coli* O antigen. The inventors now believe that many of the nucleic acid molecules derived from sugar pathway genes expressed in *S. enterica* or other enterobacteria are also likely to lack the necessary nucleotide sequence 30 specificity to identify specific O antigens or specific serotypes.

In this regard it is important to note that the genes for the synthesis of a polysaccharide antigen include those related to the synthesis of the sugars present in 35 the antigen (sugar pathway genes) and those related to the manipulation of those sugars to form the polysaccharide. The present invention is predominantly concerned with the latter group of genes, particularly the assembly and

transport genes such as transferase, polymerase and flippase genes.

The present inventors have surprisingly found that the use of nucleic acid molecules derived from particular assembly and transport genes, particularly transferase, *wzx* and *wzy* genes, within O antigen gene clusters can improve the specificity of the detection and identification of O antigens. The present inventors believe that the invention is not necessarily limited to the detection of the particular O antigens which are encoded by the nucleic acid molecules exemplified herein, but has broad application for the detection of bacteria which express an O antigen and the identification of O antigens in general. Further because of the similarities between the gene clusters involved in the synthesis of O antigens and other polymorphic polysaccharide antigens, such as bacterial capsular antigens, the inventors believe that the methods and molecules of the present invention are also applicable to these other polysaccharide antigens.

Accordingly, in one aspect the present invention relates to the identification of nucleic acid molecules which are useful for the detection and identification of specific bacterial polysaccharide antigens.

The invention provides a nucleic acid molecule derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit, including a *wzx* gene, *wzy* gene, or a gene with a similar function; the gene being involved in the synthesis of a particular bacterial polysaccharide antigen, wherein the sequence of the nucleic acid molecule is specific to the particular bacterial polysaccharide antigen.

Polysaccharide antigens, such as capsular antigens of *E. coli* (Type I and Type II), the Virulence capsule of *S. enterica* sv Typhi and the capsules of species such as *Streptococcus pneumoniae* and *Staphylococcus albus* are

encoded by genes which include nucleotide sugar pathway genes, sugar transferase genes and genes for the transport and processing of the polysaccharide or oligosaccharide unit. In some cases these are *wzx* or *wzy* but in other 5 cases they are quite different because a different processing pathway is used. Examples of other gene clusters include the gene clusters for an extracellular polysaccharide of *Streptococcus thermophilus*, an exopolysaccharide of *Rhizobium meliloti* and the K2 10 capsule of *Klebsiella pneumoniae*. These all have genes which by experimental analysis, comparison of nucleotide sequence or predicted protein structure, can be seen to include nucleotide sugar pathway genes, sugar transferase genes and genes for oligosaccharide or polysaccharide 15 processing.

In the case of the *E. coli* K-12 colanic acid capsule gene cluster [Stevenson et al (1996) "Organization of the *Escherichia coli* K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid". J. Bacteriol **178**: 4885-4893] genes from the three 20 classes were identified either provisionally or definitively. Colanic acid capsule is classified with the Type I capsule of *E. coli*.

The present inventors believe that, in general, 25 transferase genes and genes for oligosaccharide processing will be more specific for a given capsule than the genes coding for the nucleotide sugar synthetic pathways as most sugars present in such capsules occur in the capsules of different serotypes. Thus the nucleotide sugar synthesis 30 pathway genes could now be predicted to be common to more than one capsule type.

As elaborated below the present inventors recognise 35 that there may be polysaccharide antigen gene clusters which share transferase genes and/or genes for oligosaccharide or polysaccharide processing so that completely random selection of nucleotide sequences from within these genes may still lead to cross-reaction; an example with respect to capsular antigens is provided by

the *E. coli* type II capsules for which only transferase genes are sufficiently specific. However, the present inventors in light of their current results nonetheless consider the transferase genes or genes controlling oligosaccharide or polysaccharide processing to be superior targets for nucleotide sequence selection for the specific detection and characterisation of polysaccharide antigen types. Thus where there is similarity between particular genes, selection of nucleotide sequences from within other transferase genes or genes for oligosaccharide or polysaccharide processing from within the relevant gene cluster will still provide specificity, or alternatively the use of combinations of nucleotide sequences will provide the desired specificity. The combinations of nucleotide sequences may include nucleotide sequences derived from pathway genes together with nucleotide sequences derived from transferase, *wzx* or *wzy* genes.

Thus the invention also provides a panel of nucleic acid molecules wherein the nucleic acid molecules are derived from a combination of genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes; wherein the combination of genes is specific to the synthesis of a particular bacterial polysaccharide antigen and wherein the panel of nucleic acid molecules is specific to a bacterial polysaccharide antigen. In another preferred form, the nucleic acid molecules are derived from a combination of genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, together with nucleic acid molecules derived from pathway genes.

In a second aspect the present invention relates to the identification of nucleic acid molecules which are useful for the detection of bacteria which express O antigens and for the identification of the O antigens of those bacteria in diagnostic assays.

5 The invention provides a nucleic acid molecule derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit such as a *wzx* or *wzy* gene, the gene being involved in the synthesis of a particular bacterial O antigen, wherein the sequence of the nucleic acid molecule is specific to the particular bacterial O antigen.

10 The nucleic acids of the invention may be variable in length. In one embodiment they are from about 10 to about 20 nucleotides in length.

15 In one preferred embodiment, the invention provides a nucleic acid molecule derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit including a *wzx* or *wzy* gene the gene being involved in the synthesis of an O antigen expressed by *E. coli*, wherein the sequence of the nucleic acid molecule is specific to the O antigen.

20 In one more preferred embodiment, the sequence of the nucleic acid molecule is specific to the nucleotide sequence encoding the O111 antigen (SEQ ID NO:1). More preferably, the sequence is derived from a gene selected from the group consisting of *wbdH* (nucleotide position 739 to 1932 of SEQ ID NO:1), *wzx* (nucleotide position 8646 to 9911 of SEQ ID NO:1), *wzy* (nucleotide position 9901 to 10953 of SEQ ID NO:1), *wbdM* (nucleotide position 11821 to 12945 of SEQ ID NO:1) and fragments of those molecules of at least 10-12 nucleotides in length. Particularly preferred nucleic acid molecules are those set out in Table 5 and 5A, with respect to the above mentioned genes.

30 In another more preferred embodiment, the sequence of the nucleic acid molecule is specific to the nucleotide sequence encoding the O157 antigen (SEQ ID NO:2). More preferably the sequence is derived from a gene selected from the group consisting of *wbdN* (nucleotide position 79 to 861 of SEQ ID NO:2), *wbdO*, (nucleotide position 2011 to 2757 of SEQ ID NO:2), *wbdP* (nucleotide position 5257 to

6471 of SEQ ID NO:2)), *wbdR* (13156 to 13821 of SEQ ID NO:2), *wzx* (nucleotide position 2744 to 4135 of SEQ ID NO:2) and *wzy* (nucleotide position 858 to 2042 of SEQ ID NO:2). Particularly preferred nucleic acid molecules are 5 those set out in Table 6 and 6A.

The invention also provides in a further preferred embodiment a nucleic acid molecule derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or 10 oligosaccharide unit including a *wzx* or *wzy* gene; the gene being involved in the synthesis of an O antigen expressed by Salmonella enterica, wherein the sequence of the nucleic acid molecule is specific to the O antigen.

In one more preferred form of this embodiment, the 15 sequence of the nucleic acid molecule is specific to the nucleotide sequence encoding the S. enterica C2 antigen (SEQ ID NO:3). More preferably the sequence of the nucleic acid molecule is derived from a gene selected from the group consisting of *wbaR* (nucleotide position 2352 to 20 3314 of SEQ ID NO:3), *wbaL* (nucleotide position 3361 to 3875 of SEQ ID NO:3), *wbaQ* (nucleotide position 3977 to 5020 of SEQ ID NO:3), *wbaW* (nucleotide position 6313 to 7323 of SEQ ID NO:3), *wbaZ* (nucleotide position 7310 to 8467 of SEQ ID NO:3), *wzx* (nucleotide position 1019 to 25 2359 of SEQ ID NO:3) and *wzy* (nucleotide position 5114 to 6313 of SEQ ID NO:3). Particularly preferred nucleic acid molecules are those set out in Table 7.

In another more preferred form of this embodiment, 30 the sequence of the nucleic acid molecule is specific to the nucleotide sequence encoding the S. enterica B antigen (SEQ ID NO:4). More preferably the sequence is derived from *wzx* (nucleotide position 12762 to 14054 of SEQ ID NO:4) or *wbaV* (nucleotide position 14059 to 15060 of SEQ ID NO:4). Particularly preferred nucleic acid molecules 35 are those set out in Table 8 which are derived from *wzx* and *wbaV* genes.

In a further more preferred form of this embodiment, the sequence of the nucleic acid molecule is specific to

the S. enterica D3 O antigen and is derived from the *wzy* gene.

5 In yet a further preferred form of this embodiment, the sequence of the nucleic acid molecule is specific to the S. enterica E1 O antigen and is derived from the *wzx* gene.

10 While transferase genes, or genes coding for the transport or processing of a polysaccharide or oligosaccharide unit, such as a *wzx* or *wzy* gene, are superior targets for specific detection of individual O antigen types there may well be individual genes or parts of them within this group that can be demonstrated to be the same or closely related between different O antigen types such that cross-reactions can occur. Cross 15 reactions should be avoided by the selection of a different target within the group or the use of multiple targets within the group.

20 Further, it is recognised that there are cases where O antigen gene clusters have arisen from recombination of at least two strains such that the unique O antigen type is provided by a combination of gene products shared with at least two other O antigen types. The recognised example of this phenomenon is the S. enterica O antigen serotype D2 which has genes from D1 and E1 but none unique 25 to D2. In these circumstances the detection of the O antigen type can still be achieved in accordance with the invention, but requires the use of a combination of nucleic acid molecules to detect a specific combination of genes that exists only in that particular O antigen gene 30 cluster.

Thus, the invention also provides a panel of nucleic acid molecules wherein the nucleic acid molecules are derived from genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or 35 oligosaccharide unit including *wzx* or *wzy* genes, wherein the panel of nucleic acid molecules is specific to a bacterial O antigen. Preferably the particular bacterial O antigen is expressed by S. enterica. More preferably,

the panel of nucleic acid molecules is specific to the D2 O antigen and is derived from the E1 *wzy* gene and the D1 *wzx* gene.

5 The combinations of nucleotide sequences may include nucleotide sequences derived from pathway genes, together with nucleotide sequences derived from transferase, *wzx* or *wzy* genes.

10 Thus, the invention also provides a panel of nucleic acid molecules, wherein the nucleic acid molecules are derived from genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, and sugar pathway genes, wherein the panel of nucleic acid molecules is specific to a particular bacterial O antigen.

15 Preferably the O antigen is expressed *S. enterica*.

20 Further it is recognised that there may be instances where spurious hybridisation will arise through initial selection of a sequence found in many different genes but this is typically recognisable by, for instance, comparison of band sizes against controls in PCR gels, and an alternative sequence can be selected.

25 The present inventors believe that based on the teachings of the present invention and available information concerning polysaccharide antigen gene clusters (including O antigen gene clusters), and through use of experimental analysis, comparison of nucleic acid sequences or predicted protein structures, nucleic acid molecules in accordance with the invention can be readily derived for any particular polysaccharide antigen of interest. Suitable bacterial strains can typically be 30 acquired commercially from depositary institutions.

35 As mentioned above there are currently 166 defined *E. coli* O antigens while the *S. enterica* has 46 known O antigen types [Popoff M.Y. et al (1992) "Antigenic formulas of the *Salmonella* serovars" 6th revision WHO Collaborating centre for Reference and Research on *Salmonella*, Institut Pasteur Paris France]. Many other genera of bacteria are known to have O antigens and these

include Citrobacter, Shigella, Yersinia, Plesiomonas, Vibrio and Proteus.

Samples of the 166 different E. coli O antigen serotypes are available from Statens Serum Institut, 5 Copenhagen, Denmark.

The 46 S. enterica serotypes are available from Institute of Medical and Veterinary Science, Adelaide, Australia.

In another aspect, the invention relates to a method 10 of testing a sample for the presence of one or more bacterial polysaccharide antigens comprising contacting the sample with at least one oligonucleotide molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme 15 for transport or processing of oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the bacterial polysaccharide antigen; under conditions suitable to permit the at least one oligonucleotide molecule to 20 specifically hybridise to at least one such gene of any bacteria expressing the particular bacterial polysaccharide antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules.

Where a single specific oligonucleotide molecule is 25 unavailable a combination of molecules hybridising specifically to the target region may be used. Thus the invention provides a panel of nucleic acid molecules for use in the method of testing of the invention, wherein the nucleic acid molecules are derived from genes encoding 30 transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, wherein the panel of nucleic acid molecules is specific to a particular bacterial polysaccharide. The panel of nucleic acid molecules can 35 include nucleic acid molecules derived from sugar pathway genes where necessary.

In another aspect, the invention relates to a method of testing a sample for the presence of one or more

bacterial polysaccharide antigens comprising contacting the sample with at least one pair of oligonucleotide molecules, with at least one oligonucleotide molecule of the pair capable of specifically hybridising to: (i) a 5 gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the bacterial polysaccharide antigen; under conditions suitable to 10 permit the at least one oligonucleotide molecule of the pair of molecules to specifically hybridise to at least one such gene of any bacteria expressing the particular bacterial polysaccharide antigen present in the sample and detecting any specifically hybridised oligonucleotide 15 molecules.

The pair of oligonucleotide molecules may both hybridise to the same gene or to different genes. Only one oligonucleotide molecule of the pair need hybridise specifically to sequence specific for the particular 20 antigen type. The other molecule can hybridise to a non-specific region.

Where the particular polysaccharide antigen gene cluster has arisen through recombination, the at least one pair of oligonucleotide molecules may be selected to be 25 capable of hybridising to a specific combination of genes in the cluster specific to that polysaccharide antigen, or multiple pairs may be selected to provide hybridisation to the specific combination of genes. Even where all the genes in a particular cluster are unique, the method may 30 be carried out using nucleotide molecules which recognise a combination of genes within the cluster.

Thus the invention provides a panel containing pairs of nucleic acid molecules for use in the method of testing of the invention, wherein the pairs of nucleic acid 35 molecules are derived from genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, wherein the panel of nucleic acid molecules is

specific to a particular bacterial polysaccharide antigen. The panel of nucleic acid molecules can include pairs of nucleic acid molecules derived from sugar pathway genes where necessary.

5 In another aspect, the invention relates to a method of testing a sample for the presence of one or more particular bacterial O antigens comprising contacting the sample with at least one oligonucleotide molecule capable of specifically hybridising to: (i) a gene encoding an O antigen transferase, or (ii) a gene encoding an enzyme for 10 transport or processing of the oligosaccharide or polysaccharide unit, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the particular O antigen; under conditions suitable to permit the at least 15 one oligonucleotide molecule to specifically hybridise to at least one such gene of any bacteria expressing the particular bacterial O antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules. Preferably the bacteria are *E. coli* or *S. 20 enterica*. More preferably, the *E. coli* express the 0157 serotype or the 0111 serotype. More preferably the *S. enterica* express the C2 or B serotype. Preferably, the method is a Southern blot method. More preferably, the nucleic acid molecule is labelled and hybridisation of the 25 nucleic acid molecule is detected by autoradiography or detection of fluorescence.

30 The inventors envisage circumstances where a single specific oligonucleotide molecule is unavailable. In these circumstances a combination of molecules hybridising specifically to the target region may be used. Thus the invention provides a panel of nucleic acid molecules for use in the method of testing of the invention, wherein the nucleic acid molecules are derived from genes encoding transferases and/or enzymes for the transport or 35 processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, wherein the panel of nucleic acid molecules is specific to a particular bacterial O antigen. Preferably the particular bacterial O antigen is

expressed by S. enterica. The panel of nucleic acid molecules can include nucleic acid molecules derived from sugar pathway genes where necessary.

In another aspect, the invention relates to a method 5 of testing a sample for the presence of one or more particular bacterial O antigens comprising contacting the sample with at least one pair of oligonucleotide molecules with at least one oligonucleotide molecule of the pair being capable of specifically hybridising to: (i) a gene 10 encoding an O antigen transferase, or (ii) a gene encoding an enzyme for transport or processing of the oligosaccharide or polysaccharide unit, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the particular O antigen; under conditions suitable to 15 permit the at least one oligonucleotide molecule to specifically hybridise to at least one such gene of any bacteria expressing the particular bacterial O antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules.

20 Preferably the bacteria are E. coli or S. enterica. More preferably, the E. coli are of the 0111 or the 0157 serotype. More preferably the S. enterica express the C2 or B serotype. Preferably, the method is a polymerase 25 chain reaction method. More preferably the oligonucleotide molecules for use in the method of the invention are labelled. Even more preferably the hybridised oligonucleotide molecules are detected by electrophoresis. Preferred oligonucleotides for use with 0111 which provide 30 for specific detection of 0111 are illustrated in Table 5 and 5A with respect to the genes *wbdH*, *wzx*, *wzy* and *wbdM*. Preferred oligonucleotide molecules for use with 0157 which provide for specific detection of 0157 are illustrated in Table 6 and 6A.

35 With respect to serotypes C2 and B, suitable oligonucleotide molecules can be selected from appropriate regions described in column 3 of Tables 7 and 8.

The inventors envisage rare circumstances whereby two genetically similar gene clusters encoding serologically

different O antigens have arisen through recombination of genes or mutation so as to generate polymorphic variants. In these circumstances multiple pairs of oligonucleotides may be selected to provide hybridisation to the specific 5 combination of genes. The invention thus provides a panel containing pairs of nucleic acid molecules for use in the method of testing of the invention, wherein the pairs of nucleic acid molecules are derived from genes encoding transferases and/or enzymes for the transport or 10 processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, wherein the panel of nucleic acid molecules is specific to a particular bacterial O antigen. Preferably the particular bacterial O antigen is expressed by S. enterica. The panel of nucleic acid 15 molecules can include pairs of nucleic acid molecules derived from sugar pathway genes where necessary.

In another aspect, the invention relates to a method for testing a food derived sample for the presence of one or more particular bacterial O antigens comprising 20 contacting the sample with at least one pair of oligonucleotide molecules with at least one oligonucleotide molecule of the pair being capable of specifically hybridising to: (i) a gene encoding an O antigen transferase, or (ii) a gene encoding an enzyme for 25 transport or processing of the oligosaccharide or polysaccharide unit, including a *wzx* or *wzy* gene; wherein the gene is involved in the synthesis of the particular O antigen; under conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to 30 at least one such gene of any bacteria expressing the particular bacterial polysaccharide antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules. Preferably the bacteria are E. coli or S. enterica. More preferably, the E. coli are of 35 the 0111 or 0157 serotype. More preferably the S. enterica are of the C2 or B serotype. Preferably, the method is a polymerase chain reaction method. More preferably the oligonucleotide molecules for use in the

method of the invention are labelled. Even more preferably the hybridised oligonucleotide molecules are detected by electrophoresis.

In another aspect the present invention relates to a 5 method for testing a faecal derived sample for the presence of one or more particular bacterial O antigens comprising contacting the sample with at least one pair of oligonucleotide molecules with at least one oligonucleotide molecule of the pair being capable of specifically 10 hybridising to: (i) a gene encoding an O antigen transferase, or (ii) a gene encoding an enzyme for transport or processing of the oligosaccharide or polysaccharide unit, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the particular O 15 antigen; under conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to at least one of said genes of any bacteria expressing the particular bacterial O antigen present in the sample and detecting any specifically hybridised oligonucleotide 20 molecules. Preferably the bacteria are *E. coli* or *S. enterica*. More preferably, the *E. coli* are of the 0111 or 0157 serotype. More preferably, the *S. enterica* are of the C2 or B serotype. Preferably, the method is a 25 polymerase chain reaction method. More preferably the oligonucleotide molecules for use in the method of the invention are labelled. Even more preferably the hybridised oligonucleotide molecules are detected by electrophoresis.

In another aspect, the present invention relates to a 30 method for testing a sample derived from a patient for the presence of one or more particular bacterial O antigens comprising contacting the sample with at least one pair of oligonucleotide molecules with at least one oligonucleotide molecule of the pair being capable of specifically 35 hybridising to: (i) a gene encoding an O antigen transferase, or (ii) a gene encoding an enzyme for transport or processing of the oligosaccharide or polysaccharide unit, including a *wzx* or *wzy* gene; wherein

5 said gene is involved in the synthesis of the particular O antigen; under conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to at least one such gene of any bacteria expressing the particular bacterial O antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules. Preferably the bacteria are E. coli or S. enterica. More preferably, the E. coli are of the 0111 or 0157 serotype. More preferably, the S. enterica are of 10 the C2 or B serotype. Preferably, the method is a polymerase chain reaction method. More preferably the oligonucleotide molecules for use in the method of the invention are labelled. Even more preferably the hybridised oligonucleotide molecules are detected by 15 electrophoresis.

20 In the above described methods it will be understood that where pairs of oligonucleotides are used one of the oligonucleotide sequences may hybridise to a sequence that is not from a transferase, wzx or wzy gene. Further where both hybridise to one of these gene products they may hybridise to the same or a different one of these genes.

25 In addition it will be understood that where cross reactivity is an issue a combination of oligonucleotides may be chosen to detect a combination of genes to provide specificity.

30 The invention further relates to a diagnostic kit which can be used for the detection of bacteria which express bacterial polysaccharide antigens and the identification of the bacterial polysaccharide type of those bacteria.

35 Thus in a further aspect, the invention relates to a kit comprising a first vial containing a first nucleic acid molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide, including a wzx or wzy gene, wherein the said gene is involved in the synthesis of a bacterial polysaccharide. The kit may also provide in the same or a

separate vial a second specific nucleic acid capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide, 5 including a *wzx* or *wzy* gene, wherein the said gene is involved in the synthesis of a bacterial polysaccharide, wherein the sequence of the second nucleic acid molecule is different from the sequence of the first nucleic acid molecule.

10 In a further aspect the invention relates to a kit comprising a first vial containing a first nucleic acid molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide including *wzx* or *wzy*, wherein the said gene is involved in the synthesis of a bacterial O antigen. 15 The kit may also provide in the same or a separate vial a second specific nucleic acid capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide including *wzx* or *wzy*, wherein the said gene is involved in the synthesis of O antigen, wherein the sequence of the second nucleic acid molecule is different from the sequence of the first 20 nucleic acid molecule. Preferably the first and second nucleic acid sequences are derived from *E. coli* or the first and second nucleic acid sequences are derived from *S. enterica*. 25

30 The present inventors provide full length sequence of the O157 gene cluster for the first time and recognise that from this sequence of this previously uncloned full gene cluster appropriate recombinant molecules can be generated and inserted for expression to provide expressed O157 antigens useful in applications such as vaccines.

35

DEFINITIONS

The phrase, "a nucleic acid molecule derived from a gene" means that the nucleic acid molecule has a

nucleotide sequence which is either identical or substantially similar to all or part of the identified gene. Thus a nucleic acid molecule derived from a gene can be a molecule which is isolated from the identified 5 gene by physical separation from that gene, or a molecule which is artificially synthesised and has a nucleotide sequence which is either identical to or substantially similar to all or part of the identified gene. While some workers consider only the DNA strand with the same 10 sequence as the mRNA transcribed from the gene, here either strand is intended.

Transferase genes are regions of nucleic acid which have a nucleotide sequence which encodes gene products that transfer monomeric sugar units.

15 Flippase or *wzx* genes are regions of nucleic acid which have a nucleotide sequence which encodes a gene product that flips oligosaccharide repeat units generally composed of three to six monomeric sugar units to the external surface of the membrane.

20 Polymerase or *wzy* genes are regions of nucleic acid which have a nucleotide sequence which encodes gene products that polymerise repeating oligosaccharide units generally composed of 3-6 monomeric sugar units.

25 The nucleotide sequences provided in this specification are described in the sequence listing as anti-sense sequences. This term is used in the same manner as it is used in *Glossary of Biochemistry and Molecular Biology Revised Edition*, David M. Glick, 1997 Portland Press Ltd., London on page 11 where the term is 30 described as referring to one of the two strands of double-stranded DNA usually that which has the same sequence as the mRNA. We use it to describe this strand which has the same sequence as the mRNA.

NOMENCLATURESynonyms for E. coli O111 rfb

	<u>Current names</u>	<u>Our names</u>	<u>Bastin et al. 1991</u>
	wbdH	orf1	
5	gmd	orf2	
	wbdI	orf3	orf3.4*
	manC	orf4	rfbM*
	manB	orf5	rfbK*
10	wbdJ	orf6	orf6.7*
	wbdK	orf7	orf7.7*
	wzx	orf8	orf8.9 and rfbX*
	wzy	orf9	
	wbdL	orf10	
	wbdM	orf11	

15 * Nomenclature according to Bastin D.A., et al. 1991 "Molecular cloning and expression in Escherichia coli K-12 of the rfb gene cluster determining the O antigen of an E. coli O111 strain". *Mol. Microbiol.* 5:9 2223-2231.

20 Other Synonyms

	wzy	rfc	
	wzx	rfbX	
	rmlA	rfbA	
	rmlB	rfbB	
25	rmlC	rfbC	
	rmlD	rfbD	
	glf	orf6*	
	wbbI	orf3#, orf8* of <u>E. coli</u> K-12	
	wbbJ	orf2#, orf9* of <u>E. coli</u> K-12	
30	wbbK	orf1#, orf10* of <u>E. coli</u> K-12	
	wbbL	orf5#, orf 11* of <u>E. coli</u> K-12	

Nomenclature according to Yao, Z. And M. A. Valvano 1994. "Genetic analysis of the O-specific lipopolysaccharide biosynthesis region (rfb) of Escherichia coli K-12 W3110: identification of genes the confer groups-specificity to Shigella flexineri serotypes Y and 4a". *J. Bacteriol.* 176: 4133-4143.

* Nomenclature according to Stevenson et al. 1994. "Structure of the O-antigen of E. coli K-12 and the sequence of its rfb gene cluster". *J. Bacteriol.* 176: 4144-4156.

40 • S. enterica is a name introduced in 1987 to replace the many other names such as Salmonella typhi and Salmonella typhimurium, the old species names becoming serovar names as in S. enterica sv Typhi. However, the traditional names are still widely used.

45 • The O antigen genes of many species were given rfb names (rfbA etc) and the O antigen gene cluster was often referred to as the rfb cluster. There are now new names for the rfb genes as shown in the table. Both terminologies have been used herein, depending on the source of the information.

- BRIEF DESCRIPTION OF DRAWINGS

Figure 1 shows *Eco* R1 restriction maps of cosmid clones pPR1054, pPR1055, pPR1056, pPR1058, pPR1287 which are subclones of *E. coli* O111 O antigen gene cluster. The 5 thickened line is the region common to all clones. Broken lines show segments that are non-contiguous on the chromosome. The deduced restriction map for *E. coli* strain M92 is shown above.

Figure 2 shows a restriction mapping analysis of *E. coli* O111 O antigen gene cluster within the cosmid clone 10 pPR1058. Restriction enzymes are: (B: *Bam*H1; Bg: *Bgl*III, E: *Eco*R1; H: *Hind*III; K: *Kpn*I; P: *Pst*I; S: *Sal*I and X: *Xba*I. Plasmids pPR1230, pPR1231, and pPR1288 are deletion 15 derivatives of pPR1058. Plasmids pPR 1237, pPR1238, pPR1239 and pPR1240 are in pUC19. Plasmids pPR1243, pPR1244, pPR1245, pPR1246 and pPR1248 are in pUC18, and pPR1292 is in pUC19. Plasmid pPR1270 is in pT7T319U. Probes 1, 2 and 3 were isolated as internal fragments of 20 pPR1246, pPR1243 and pPR1237 respectively. Dotted lines indicate that subclone DNA extends to the left of the map into attached vector.

Figure 3 shows the structure of *E. coli* O111 O antigen gene cluster.

Figure 4 shows the structure of *E. coli* O157 O antigen gene cluster. 25

Figure 5 shows the structure *S. enterica* locus encoding the serogroup C2 O antigen gene cluster.

Figure 6 shows the structure *S. enterica* locus encoding the serogroup B O antigen gene cluster.

30 Figure 7 shows the nucleotide sequence of the *E. coli* O111 O antigen gene cluster. Note: (1) The first and last three bases of a gene are underlined and of italic respectively.; (2) The region which was previously sequenced by Bastin and Reeves 1995 "Sequence and analysis 35 of the O antigen gene (rfb) cluster of *Escherichia coli* o111" Gene 164: 17-23 is marked.

Figure 8 shows the nucleotide sequence of the *E. coli* O157 O antigen gene cluster. Note: (1) The first and last

three bases of a gene (region) are underlined and of italic respectively (2) The region previously sequenced by Bilge et al. 1996 "Role of the Escherichia coli O157-H7 O side chain in adherence and analysis of an rfb locus". Inf. and 5 Immun 64:4795-4801 is marked.

Figure 9 shows the nucleotide sequence of S. enterica serogroup C2 O antigen gene cluster. Note:

(1) The numbering is as in Brown et al. 1992. "Molecular analysis of the rfb gene cluster of *Salmonella* serovar 10 muenchen (strain M67): the genetic basis of the polymorphism between groups C2 and B". Mol. Microbiol. 6: 1385-1394 (2) The first and last three bases of a gene are underlined and in italics respectively. (3) Only that part 15 of the group C2 gene cluster, which differs from that of group B, was sequenced and is presented here.

Figure 10 shows the nucleotide sequence of S. enterica serogroup B O antigen gene cluster Note: (1) The numbering is as in Jiang et al. 1991. "Structure and sequence of the rfb (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain 20 LT2)". Mol. Microbiol. 5: 695-713. The first gene in the O antigen gene cluster is *rmlB* which starts at base 4099. (2) The first and last three bases of a gene are underlined and in italics respectively.

25 **BEST METHOD FOR CARRYING OUT THE INVENTION**

Materials and Methods-part 1

The experimental procedures for the isolation and characterisation of the E. coli O111 O antigen gene cluster (position 3,021-9,981) are according to Bastin 30 D.A., et al. 1991 "Molecular cloning and expression in Escherichia coli K-12 of the rfb gene cluster determining the O antigen of an E. coli O111 strain". Mol. Microbiol. 5:9 2223-2231 and Bastin D.A. and Reeves, P.R. 1995 "Sequence and analysis of the O antigen gene(rfb)cluster 35 of Escherichia coli O111". Gene 164: 17-23.

A. Bacterial strains and growth media

Bacteria were grown in Luria broth supplemented as required.

B. Cosmids and phage

Cosmids in the host strain x2819 were repackaged *in vivo*. Cells were grown in 250mL flasks containing 30mL of culture, with moderate shaking at 30°C to an optical density of 0.3 at 580 nm. The defective lambda prophage was induced by heating in a water bath at 45°C for 15min followed by an incubation at 37°C with vigorous shaking for 2hr. Cells were then lysed by the addition of 0.3mL chloroform and shaking for a further 10min. Cell debris were removed from 1mL of lysate by a 5min spin in a microcentrifuge, and the supernatant removed to a fresh microfuge tube. One drop of chloroform was added then shaken vigorously through the tube contents.

C. DNA preparation

Chromosomal DNA was prepared from bacteria grown overnight at 37°C in a volume of 30mL of Luria broth. After harvesting by centrifugation, cells were washed and resuspended in 10mL of 50mMTris-HCl pH 8.0. EDTA was added and the mixture incubated for 20min. Then lysozyme was added and incubation continued for a further 10min. Proteinase K, SDS, and ribonuclease were then added and the mixture incubated for up to 2hr for lysis to occur. All incubations were at 37°C. The mixture was then heated to 65°C and extracted once with 8mL of phenol at the same temperature. The mixture was extracted once with 5mL of phenol/chloroform/iso-amyl alcohol at 4°C. Residual phenol was removed by two ether extractions. DNA was precipitated with 2 vols. of ethanol at 4°C, spooled and washed in 70% ethanol, resuspended in 1-2mL of TE and dialysed. Plasmid and cosmid DNA was prepared by a modification of the Birnboim and Doly method [Birnboim, H. C. And Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA *Nucl. Acid Res.* 7:1513-1523. The volume of culture was 10mL and the lysate was extracted with phenol/chloroform/iso-amyl alcohol before precipitation with isopropanol. Plasmid

- 27 -

DNA to be used as vector was isolated on a continuous caesium chloride gradient following alkaline lysis of cells grown in 1L of culture.

D. Enzymes and buffers.

5 Restriction endonucleases and DNA T4 ligase were purchased from Boehringer Mannheim (Castle Hill, NSW, Australia) or Pharmacia LKB (Melbourne, VIC Australia). Restriction enzymes were used in the recommended commercial buffer.

10 E. Construction of a gene bank.

Individual aliquots of M92 chromosomal DNA (strain Stoke W, from Statens Serum Institut, 5 Artillerivej, 2300 Copenhagen S, Denmark) were partially digested with 0.2U Sau3A1 for 1-15mins. Aliquots giving the greatest proportion of fragments in the size range of approximately 15 40-50kb were selected and ligated to vector pPR691 previously digested with *Bam*H1 and *Pvu*II. Ligation mixtures were packaged *in vitro* with packaging extract. The host strain for transduction was x2819 and 20 recombinants were selected with kanamycin.

F. Serological procedures.

Colonies were screened for the presence of the O111 antigen by immunoblotting. Colonies were grown overnight, up to 100 per plate then transferred to nitrocellulose 25 discs and lysed with 0.5N HCl. Tween 20 was added to TBS at 0.05% final concentration for blocking, incubating and washing steps. Primary antibody was *E. coli* O group 111 antiserum, diluted 1:800. The secondary antibody was goat anti-rabbit IgG labelled with horseradish peroxidase 30 diluted 1:5000. The staining substrate was 4-chloro-1-naphthol. Slide agglutination was performed according to the standard procedure.

G. Recombinant DNA methods.

35 Restriction mapping was based on a combination of standard methods including single and double digests and sub-cloning. Deletion derivatives of entire cosmids were produced as follows: aliquots of 1.8 μ g of cosmid DNA were

digested in a volume of 20 μ l with 0.25U of restriction enzyme for 5-80min. One half of each aliquot was used to check the degree of digestion on an agarose gel. The sample which appeared to give a representative range of 5 fragments was ligated at 4°C overnight and transformed by the CaCl₂ method into JM109. Selected plasmids were transformed into *λ*174 by the same method. P4657 was transformed with pPR1244 by electroporation.

H. DNA hybridisation

10 Probe DNA was extracted from agarose gels by electroelution and was nick-translated using [α -32P]-dCTP. Chromosomal or plasmid DNA was electrophoresed in 0.8% agarose and transferred to a nitrocellulose membrane. The hybridisation and pre-hybridisation buffers contained 15 either 30% or 50% formamide for low and high stringency probing respectively. Incubation temperatures were 42°C and 37°C for pre-hybridisation and hybridisation respectively. Low stringency washing of filters consisted of 3 x 20min washes in 2 x SSC and 0.1% SDS. High- 20 stringency washing consisted of 3 x 5min washes in 2 x SSC and 0.1% SDS at room temperature, a 1hr wash in 1 x SSC and 0.1% SDS at 58°C and 15min wash in 0.1 x SSC and 0.1% SDS at 58°C.

I. Nucleotide sequencing of *E. coli* O111 O antigen gene 25 cluster (position 3,021-9,981)

Nucleotide sequencing was performed using an ABI 373 automated sequencer (CA, USA). The region between map positions 3.30 and 7.90 was sequenced using uni-directional exonuclease III digestion of deletion 30 families made in PT7T3190 from clones pPR1270 and pPR1272. Gaps were filled largely by cloning of selected fragments into M13mp18 or M13mp19. The region from map positions 7.90-10.2 was sequenced from restriction fragments in M13mp18 or M13mp19. Remaining gaps in both the regions 35 were filled by priming from synthetic oligonucleotides complementary to determined positions along the sequence,

using a single stranded DNA template in M13 or phagemid. The oligonucleotides were designed after analysing the adjacent sequence. All sequencing was performed by the chain termination method. Sequences were aligned using SAP [Staden, R., 1982 "Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing". *Nuc. Acid Res.* 10: 4731-4751; Staden, R., 1986 "The current status and portability of our sequence handling software". *Nuc. Acid Res.* 14: 217-231]. The program NIP [Staden, R. 1982 "An interactive graphics program for comparing and aligning nucleic acid and amino acid sequence". *Nuc. Acid Res.* 10: 2951-2961] was used to find open reading frames and translate them into proteins.

J. Isolation of clones carrying E. coli 0111 O antigen gene cluster

The E. coli O antigen gene cluster was isolated according to the method of Bastin D.A., et al. [1991 "Molecular cloning and expression in *Escherichia coli* K-12 of the *rfb* gene cluster determining the O antigen of an E. coli 0111 strain". *Mol. Microbiol.* 5(9), 2223-2231]. Cosmid gene banks of M92 chromosomal DNA were established in the *in vivo* packaging strain x2819. From the genomic bank, 3.3×10^3 colonies were screened with E. coli 0111 antiserum using an immuno-blotting procedure: 5 colonies (pPR1054, pPR1055, pPR1056, pPR1058 and pPR1287) were positive. The cosmids from these strains were packaged *in vivo* into lambda particles and transduced into the E. coli deletion mutant SΦ174 which lacks all O antigen genes. In this host strain, all plasmids gave positive agglutination with 0111 antiserum. An *Eco* R1 restriction map of the 5 independent cosmids showed that they have a region of approximately 11.5 kb in common (Figure 1). Cosmid pPR1058 included sufficient flanking DNA to identify several chromosomal markers linked to O antigen gene cluster and was selected for analysis of the O antigen gene cluster region.

K. Restriction mapping of cosmid pPR1058

- 30 -

Cosmid pPR1058 was mapped in two stages. A preliminary map was constructed first, and then the region between map positions 0.00 and 23.10 was mapped in detail, since it was shown to be sufficient for O111 antigen expression. Restriction sites for both stages are shown in Figure 2. The region common to the five cosmid clones was between map positions 1.35 and 12.95 of pPR1058.

To locate the O antigen gene cluster within pPR1058, pPR1058 cosmid was probed with DNA probes covering O antigen gene cluster flanking regions from *S. enterica* LT2 and *E. coli* K-12. Capsular polysaccharide (*cps*) genes lie upstream of O antigen gene cluster while the gluconate dehydrogenase (*gnd*) gene and the histidine (*his*) operon are downstream, the latter being further from the O antigen gene cluster. The probes used were pPR472 (3.35kb), carrying the *gnd* gene of LT2, pPR685 (5.3kb) carrying two genes of the *cps* cluster, *cpsB* and *cpsG* of LT2, and K350 (16.5kb) carrying all of the *his* operon of K-12. Probes hybridised as follows: pPR472 hybridised to 1.55kb and 3.5 kb (including 2.7 kb of vector) fragments of *Pst*1 and *Hind*III double digests of pPR1246 (a *Hind*III/*Eco*R1 subclone derived from pPR1058, Figure 2), which could be located at map positions 12.95-15.1; pPR685 hybridised to a 4.4 kb *Eco*R1 fragment of pPR1058 (including 1.3 kb of vector) located at map position 0.00-3.05; and K350 hybridised with a 32kb *Eco*R1 fragment of pPR1058 (including 4.0kb of vector), located at map position 17.30-45.90. Subclones containing the presumed *gnd* region complemented a *gnd*⁻*edd*⁻ strain GB23152. On gluconate bromothymol blue plates, pPR1244 and pPR1292 in this host strain gave the green colonies expected of a *gnd*⁻*edd*⁻ genotype. The *his*⁺ phenotype was restored by plasmid pPR1058 in the *his* deletion strain S ϕ 174 on minimal medium plates, showing that the plasmid carries the entire *his* operon.

It is likely that the O antigen gene cluster region lies between *gnd* and *cps*, as in other *E. coli* and *S. enterica* strains, and hence between the approximate map

positions 3.05 and 12.95. To confirm this, deletion derivatives of pPR1058 were made as follows: first, pPR1058 was partially digested with *Hind*III and self ligated. Transformants were selected for kanamycin 5 resistance and screened for expression of O111 antigen. Two colonies gave a positive reaction. *Eco*R1 digestion showed that the two colonies hosted identical plasmids, one of which was designated pPR1230, with an insert which extended from map positions 0.00 to 23.10. Second pPR1058 10 was digested with *Sall* and partially digested with *Xho*1 and the compatible ends were re-ligated. Transformants were selected with kanamycin and screened for O111 antigen expression. Plasmid DNA of 8 positively reacting clones was checked using *Eco*R1 and *Xho*1 digestion and appeared to 15 be identical. The cosmid of one was designated pPR1231. The insert of pPR1231 contained the DNA region between map positions 0.00 and 15.10. Third, pPR1231 was partially digested with *Xho*1, self-ligated, and transformants selected on spectinomycin/ streptomycin plates. Clones 20 were screened for kanamycin sensitivity and of 10 selected, all had the DNA region from the *Xho*1 site in the vector to the *Xho*1 site at position 4.00 deleted. These clones did not express the O111 antigen, showing that the *Xho*1 site at position 4.00 is within the O antigen gene 25 cluster. One clone was selected and named pPR1288. Plasmids pPR1230, pPR1231, and pPR1288 are shown in Figure 2.

L. Analysis of the *E. coli* O111 O antigen gene cluster (position 3,021-9,981) nucleotide sequence data

30 Bastin and Reeves [1995 "Sequence and analysis of the O antigen gene(*rfb*)cluster of *Escherichia coli* O111". Gene 164: 17-23] partially characterised the *E. coli* O111 O antigen gene cluster by sequencing a fragment from map position 3,021-9,981. Figure 3 shows the gene 35 organisation of position 3,021-9,981 of *E. coli* O111 O antigen gene cluster. *orf3* and *orf6* have high level amino acid identity with *wcaH* and *wcaG* (46.3% and 37.2% respectively), and are likely to be similar in function to

- 32 -

sugar biosynthetic pathway genes in the *E. coli* K-12 colanic gene cluster. *orf4* and *orf5* show high levels of amino acid homology to *manC* and *manB* genes respectively. *orf7* shows high level homology with *rfbH* which is an 5 abequose pathway gene. *orf8* encodes a protein with 12 transmembrane segments and has similarity in secondary structure to other *wzx* genes and is likely therefore to be the O antigen flippase gene.

10 Materials and Methods-part 2

A. Nucleotide sequencing of 1 to 3,020 and 9,982 to 14,516 of the *E. coli* O111 O antigen gene cluster

The sub clones which contained novel nucleotide sequences, pPR1231 (map position 0 and 1,510), pPR1237 (map position -300 to 2,744), pPR1239 (map position 2,744 to 4,168), pPR1245 (map position 9,736 to 12,007) and pPR1246 (map position 12,007 to 15,300) (Figure 2), were characterised as follows: the distal ends of the inserts of pPR1237, pPR1239 and pPR1245 were sequenced using the 15 M13 forward and reverse primers located in the vector. 20 PCR walking was carried out to sequence further into each insert using primers based on the sequence data and the primers were tagged with M13 forward or reverse primer sequences for sequencing. This PCR walking procedure was 25 repeated until the entire insert was sequenced. pPR1246 was characterised from position 12,007 to 14,516. The DNA of these sub clones was sequenced in both directions. The sequencing reactions were performed using the dideoxy 30 termination method and thermocycling and reaction products were analysed using fluorescent dye and an ABI automated sequencer (CA, USA).

B. Analysis of the *E. coli* O111 O antigen gene cluster (positions 1 to 3,020 and 9,982 to 14,516 of SEQ ID NO:1) nucleotide sequence data

35 The gene organisation of regions of *E. coli* O111 O antigen gene cluster which were not characterised by Bastin and Reeves [1995 "Sequence and analysis of the O antigen gene(*rfb*)cluster of *Escherichia coli* O111." Gene

164: 17-23], (positions 1 to 3,020 and 9,982 to 14,516) is shown in Figure 3. There are two open reading frames in region 1. Four open reading frames are predicted in region 2. The position of each gene is listed in Table 5.

5 The deduced amino acid sequence of *orf1* (*wbdH*) shares about 64% similarity with that of the *rfp* gene of *Shigella dysenteriae*. *Rfp* and *WbdH* have very similar hydrophobicity plots and both have a very convincing predicted transmembrane segment in a corresponding 10 position. *rfp* is a galactosyl transferase involved in the synthesis of LPS core, thus *wbdH* is likely to be a galactosyl transferase gene. *orf2* has 85.7% identity at amino acid level to the *gmd* gene identified in the *E. coli* K-12 colanic acid gene cluster and is likely to be a *gmd* 15 gene. *orf9* encodes a protein with 10 predicted transmembrane segments and a large cytoplasmic loop. This inner membrane topology is a characteristic feature of all known O antigen polymerases thus it is likely that *orf9* encodes an O antigen polymerase gene, *wzy*. *orf10* 20 (*wbdL*) has a deduced amino acid sequence with low homology with *Lsi2* of *Neisseria gonorrhoeae*. *Lsi2* is responsible for adding GlcNAc to galactose in the synthesis of lipooligosaccharide. Thus it is likely that *wbdL* is either a colitose or glucose transferase gene. *orf11* 25 (*wbdM*) shares high level nucleotide and amino acid similarity with *TrsE* of *Yersinia enterocholitica*. *TrsE* is a putative sugar transferase thus it is likely that *wbdM* encodes the colitose or glucose transferase.

In summary three putative transferase genes and an O 30 antigen polymerase gene were identified at map position 1 to 3,020 and 9,982 to 14,516 of *E. coli* O111 O antigen gene cluster. A search of GenBank has shown that there are no genes with significant similarity at the nucleotide sequence level for two of the three putative transferase 35 genes or the polymerase gene. SEQ ID NO:1 and Figure 7 provide the nucleotide sequence of the O111 antigen gene cluster.

Materials and Methods-part 3

A. PCR amplification of O157 antigen gene cluster from an *E. coli* O157:H7 strain (Strain C664-1992, from Statens Serum Institut, 5 Artillerivej, 2300, Copenhagen S,

5 Denmark)

E. coli O157 O antigen gene cluster was amplified by using long PCR [Cheng et al. 1994, Effective amplification of long targets from cloned inserts and human and genomic DNA" P.N.A.S. USA 91: 5695-569] with one primer (primer 10 #412: att ggt agc tgt aag cca agg gcg gta gcg t) based on the JumpStart sequence usually found in the promoter region of O antigen gene clusters [Hobbs, et al. 1994 "The JumpStart sequence: a 39 bp element common to several 15 polysaccharide gene clusters" Mol. Microbiol. 12: 855-856], and another primer #482 (cac tgc cat acc gac gac gcc gat 20 ctg ttg ctt gg) based on the *gnd* gene usually found downstream of the O antigen gene cluster. Long PCR was carried out using the Expand Long Template PCR System from 25 Boehringer Mannheim (Castle Hill NSW Australia), and products, 14 kb in length, from several reactions were combined and purified using the Promega Wizard PCR preps DNA purification System (Madison WI USA). The PCR product was then extracted with phenol and twice with ether, precipitated with 70% ethanol, and resuspended in 40 μ L of 30 water.

25 B. Construction of a random DNase I bank:

Two aliquots containing about 150ng of DNA each were subjected to DNase I digestion using the Novagen DNase I Shotgun Cleavage (Madison WI USA) with a modified protocol 30 as described. Each aliquot was diluted into 45 μ l of 0.05M Tris -HCl (pH7.5), 0.05mg/mL BSA and 10mM MnCl₂. 5 μ L of 1:3000 or 1:4500 dilution of DNaseI (Novagen) (Madison WI USA) in the same buffer was added into each tube 35 respectively and 10 μ l of stop buffer (100mM EDTA), 30% glycerol, 0.5% Orange G, 0.075% xylene and cyanol (Novagen) (Madison WI USA) was added after incubation at 15°C for 5 min. The DNA from the two DNaseI reaction

tubes were then combined and fractionated on a 0.8% LMT agarose gel, and the gel segment with DNA of about 1kb in size (about 1.5mL agarose) was excised. DNA was extracted from agarose using Promega Wizard PCR Preps DNA Purification (Madison WI USA) and resuspended in 200 μ L water, before being extracted with phenol and twice with ether, and precipitated. The DNA was then resuspended in 17.25 μ L water and subjected to T4 DNA polymerase repair and single dA tailing using the Novagen Single dA Tailing Kit (Madison WI USA). The reaction product (85 μ l containing about 8ng DNA) was then extracted with chloroform:isoamyl alcohol (24:1) once and ligated to 3x 10^{-3} pmol pGEM-T (Promega) (Madison WI USA) in a total volume of 100 μ L. Ligation was carried out overnight at 4°C and the ligated DNA was precipitated and resuspended in 20 μ L water before being electroporated into E. coli strain JM109 and plated out on BCIG-IPTG plates to give a bank.

C. Sequencing

DNA templates from clones of the bank were prepared for sequencing using the 96-well format plasmid DNA miniprep kit from Advanced Genetic Technologies Corp (Gaithersburg MD USA). The inserts of these clones were sequenced from one or both ends using the standard M13 sequencing primer sites located in the pGEM-T vector. Sequencing was carried out on an ABI377 automated sequencer (CA USA) as described above, after carrying out the sequencing reaction on an ABI Catalyst (CA USA). Sequence gaps and areas of inadequate coverage were PCR amplified directly from O157 chromosomal DNA using primers based on the already obtained sequencing data and sequenced using the standard M13 sequencing primer sites attached to the PCR primers.

D. Analysis of the E. coli O157 O antigen gene cluster nucleotide sequence data

Sequence data were processed and analysed using the

Staden programs [Staden, R., 1982 "Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing." *Nuc. Acid Res.* 10: 4731-4751; Staden, R., 1986 "The current status and 5 portability of our sequence handling software". *Nuc. Acid Res.* 14: 217-231; Staden, R. 1982 "An interactive graphics program for comparing and aligning nucleic acid and amino acid sequence". *Nuc. Acid Res.* 10: 2951-2961]. Figure 4 shows the structure of *E. coli* O157 O antigen gene 10 cluster. Twelve open reading frames were predicted from the sequence data, and the nucleotide and amino acid sequences of all these genes were then used to search the GenBank database for indication of possible function and specificity of these genes. The position of each gene is 15 listed in Table 6. The nucleotide sequence is presented in SEQ ID NO:2 and Figure 8.

orfs 10 and 11 showed high level identity to *manC* and *manB* and were named *manC* and *manB* respectively. *orf7* showed 89% identity (at amino acid level) to the *gmd* gene 20 of the *E. coli* colanic acid capsule gene cluster (Stevenson G., K. et al. 1996 "Organisation of the *Escherichia coli* K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid". *J. Bacteriol.* 178:4885-4893) and was named *gmd*. *orf8* showed 79% and 69% identity (at amino acid level) 25 respectively to *wcaG* of the *E. coli* colanic acid capsule gene cluster and to *wbcJ* (*orf14.8*) gene of the *Yersinia enterocolitica* O8 O antigen gene cluster (Zhang, L. et al. 1997 "Molecular and chemical characterization of the 30 lipopolysaccharide O-antigen and its role in the virulence of *Y. enterocolitica* serotype O8". *Mol. Microbiol.* 23:63-76). Colanic acid and the *Yersinia* O8 O antigen both contain fucose as does the O157 O antigen. There are two enzymatic steps required for GDP-L-fucose synthesis from 35 GDP-4-keto-6-deoxy-D-mannose, the product of the *gmd* gene product. However, it has been shown recently (Tonetti, M et al. 1996 Synthesis of GDP-L-fucose by the human FX protein *J. Biol. Chem.* 271:27274-27279) that the human FX

protein has "significant homology" with the *wcaG* gene (referred to as *Yefb* in that paper), and that the FX protein carries out both reactions to convert GDP-4-keto-6-deoxy-D-mannose to GDP-L-fucose. We believe that this 5 makes a very strong case for *orf8* carrying out these two steps and propose to name the gene *fcl*. In support of the one enzyme carrying out both functions is the observation that there are no genes other than *manB*, *manC*, *gmd* and *fcl* with similar levels of similarity between the three 10 bacterial gene clusters for fucose containing structures.

10 *orf5* is very similar to *wbeE* (*rfbE*) of *Vibrio cholerae* O1, which is thought to be the perosamine synthetase, which converts GDP-4-keto-6-deoxy-D-mannose to GDP-perosamine (Stroher, U.H et al. 1995 "A putative 15 pathway for perosamine biosynthesis is the first function encoded within the *rfb* region of *Vibrio cholerae*" O1. Gene 166: 33-42). *V. cholerae* O1 and *E. coli* O157 O antigens contain perosamine and N-acetyl-perosamine respectively. The *V. cholerae* O1 *manA*, *manB*, *gmd* and *wbeE* genes are the 20 only genes of the *V. cholerae* O1 gene cluster with significant similarity to genes of the *E. coli* O157 gene cluster and we believe that our observations both confirm the prediction made for the function of *wbe* of *V. cholerae*, and show that *orf5* of the O157 gene cluster 25 encodes GDP-perosamine synthetase. *orf5* is therefore named *per*. *orf5* plus about 100bp of the upstream region (postion 4022-5308) was previously sequenced by Bilge, S.S. et al. [1996 "Role of the *Escherichia coli* O157-H7 O side chain in adherence and analysis of an *rfb* locus". Infect. 30 Immun. 64:4795-4801].

35 *orf12* shows high level similarity to the conserved region of about 50 amino acids of various members of an acetyltransferase family (Lin, W., et al. 1994 "Sequence analysis and molecular characterisation of genes required for the biosynthesis of type 1 capsular polysaccharide in *Staphylococcus aureus*". J. Bacteriol. 176: 7005-7016) and we believe it is the N-acetyltransferase to convert GDP-perosamine to GDP-perNAc. *orf12* has been named *wbdR*.

The genes *manB*, *manC*, *gmd*, *fcl*, *per* and *wbdR* account for all of the expected biosynthetic pathway genes of the O157 gene cluster.

The remaining biosynthetic step(s) required are for synthesis of UDP-GalNAc from UDP-Glc. It has been proposed (Zhang, L., et al. 1997 "Molecular and chemical characterisation of the lipopolysaccharide O-antigen and its role in the virulence of Yersinia enterocolitica serotype O8". Mol. Microbiol. 23:63-76) that in Yersinia enterocolitica UDP-GalNAc is synthesised from UDP-GlcNAc by a homologue of galactose epimerase (GalE), for which there is a *gale* like gene in the Yersinia enterocolitica O8 gene cluster. In the case of O157 there is no *gale* homologue in the gene cluster and it is not clear how UDP-GalNAc is synthesised. It is possible that the galactose epimerase encoded by the *gale* gene in the *gal* operon, can carry out conversion of UDP-GlcNAc to UDP-GalNAc in addition to conversion of UDP-Glc to UDP-Gal. There do not appear to be any gene(s) responsible for UDP-GalNAc synthesis in the O157 gene cluster.

orf4 shows similarity to many *wzx* genes and is named *wzx* and *orf2* which shows similarity of secondary structure in the predicted protein to other *wzy* genes and is for that reason named *wzy*.

The *orf1*, *orf3* and *orf6* gene products all have characteristics of transferases, and have been named *wbdN*, *wbdO* and *wbdP* respectively. The O157 O antigen has 4 sugars and 4 transferases are expected. The first transferase to act would put a sugar phosphate onto undecaprenol phosphate. The two transferases known to perform this function, *WbaP* (*RfbP*) and *WecA* (*Rfe*) transfer galactose phosphate and N-acetyl-glucosamine phosphate respectively to undecaprenol phosphate. Neither of these sugars is present in the O157 structure.

Further, none of the presumptive transferases in the O157 gene cluster has the transmembrane segments found in *WecA* and *WbaP* which transfer a sugar phosphate to undecaprenol phosphate and expected for any protein which

transferred a sugar to undecaprenol phosphate which is embedded within the membrane.

The *WecA* gene which transfers GlcNAc-P to undecaprenol phosphate is located in the Enterobacteal 5 Common Antigen (ECA) gene cluster and it functions in ECA synthesis in most and perhaps all *E. coli* strains, and also in O antigen synthesis for those strains which have GlcNAc as the first sugar in the O unit.

It appears that *WecA* acts as the transferase for 10 addition of GalNAc-1-P to undecaprenol phosphate for the *Yersinia enterocolitica* 08 O antigen [Zhang et al. 1997 "Molecular and chemical characterisation of the 15 lipopolysaccharide O antigen and its role in the virulence of *Yersinia enterocolitica* serotype 08" Mol. Microbiol. 23: 63-76.] and perhaps does so here as the O157 structure includes GalNAc. *WecA* has also been reported to add Glucose-1-P phosphate to undecaprenol phosphate in *E. coli* 20 O8 and O9 strains, and an alternative possibility for transfer of the first sugar to undecaprenol phosphate is 25 *WecA* mediated transfer of glucose, as there is a glucose residue in the O157 O antigen. In either case the requisite number of transferase genes are present if GalNAc or Glc is transferred by *WecA* and the side chain Glc is transferred by a transferase outside of the O antigen gene cluster.

orf9 shows high level similarity (44% identity at 30 amino acid level, same length) with *wcaH* gene of the *E. coli* colanic acid capsule gene cluster. The function of this gene is unknown, and we give *orf9* the name *wbdQ*.

The DNA between *manB* and *wbdR* has strong sequence 35 similarity to one of the H-repeat units of *E. coli* K12. Both of the inverted repeat sequences flanking this region are still recognisable, each with two of the 11 bases being changed. The H-repeat associated protein encoding gene located within this region has a 267 base deletion and mutations in various positions. It seems that the H-repeat unit has been associated with this gene cluster for a long period of time since it translocated to the gene

cluster, perhaps playing a role in assembly of the gene cluster as has been proposed in other cases.

Materials and Methods - part 4

5 To test our hypothesis that O antigen genes for transferases and the *wzx*, *wzy* genes were more specific than pathway genes for diagnostic PCR, we first carried out PCR using primers for all the *E. coli* 016 O antigen genes (Table 4). The PCR was then carried out using PCR 10 primers for *E. coli* 0111 transferase, *wzx* and *wzy* genes (Table 5, 5A). PCR was also carried out using PCR primers for the *E. coli* 0157 transferase, *wzx* and *wzy* genes (Table 6, 6A).

15 Chromosomal DNA from the 166 serotypes of *E. coli* available from Statens Serum Institut, 5 Artillerivej, 2300 Copenhagen Denmark was isolated using the Promega Genomic (Madison WI USA) isolation kit. Note that 164 of the serogroups are described by Ewing W. H.: Edwards and Ewings "Identification of the Enterobacteriaceae" Elsevier, 20 Amsterdam 1986 and that they are numbered 1-171 with numbers 31, 47, 67, 72, 93, 94 and 122 no longer valid. Of the two serogroup 19 strains we used 19ab strain F8188-41. Lior H. 1994 ["Classification of *Eschericia coli* In *Eschericia coli* in domestic animals and humans pp 31-72. 25 Edited by C.L. Gyles CAB international] adds two more numbered 172 and 173 to give the 166 serogroups used. Pools containing 5 to 8 samples of DNA per pool were made. Pool numbers 1 to 19 (Table 1) were used in the *E. coli* 0111 and 0157 assay. Pool numbers 20 to 28 were also used 30 in the 0111 assay, and pool numbers 22 to 24 contained *E. coli* 0111 DNA and were used as positive controls (Table 2). Pool numbers 29 to 42 were also used in the 0157 assay, and pool numbers 31 to 36 contained *E. coli* 0157 DNA, and were used as positive controls (Table 3). Pool 35 numbers 2 to 20, 30, 43 and 44 were used in the *E. coli* 016 assay (Tables 1 to 3). Pool number 44 contained DNA of *E. coli* K-12 strains C600 and WG1 and was used as a positive control as between them they have all of the *E.*

coli K-12 O16 O antigen genes.

PCR reactions were carried out under the following conditions: denaturing 94°C/30"; annealing, temperature varies (refer to Tables 4 to 8)/30"; extension, 72°C/1'; 5 30 cycles. PCR reaction was carried out in an volume of 25µL for each pool. After the PCR reaction, 10µL PCR product from each pool was run on an agarose gel to check for amplified DNA.

10 Each E. coli and S. enterica chromosomal DNA sample was checked by gel electrophoresis for the presence of chromosomal DNA and by PCR amplification of the E. coli or S. enterica mdh gene using oligonucleotides based on E. coli K-12 or Salmonella enterica LT2 [Boyd et al. (1994) 15 "Molecular genetic basis of allelic polymorphism in malate dehydrogenase (mdh) in natural populations of *Escherichia coli* and Salmonella enterica" Proc. Nat. Acad. Sci. USA. 91:1280-1284.] Chromosomal DNA samples from other bacteria were only checked by gel electrophoresis of chromosomal DNA.

20

A. Primers based on E. coli O16 O antigen gene cluster sequence.

25 The O antigen gene cluster of E. coli O16 was the only typical E. coli O antigen gene cluster that had been fully sequenced prior to that of O111, and we chose it for testing our hypothesis. One pair of primers for each gene was tested against pools 2 to 20, 30 and 43 of E. coli chromosomal DNA. The primers, annealing temperatures and functional information for each gene are listed in Table 30 4.

35 For the five pathway genes, there were 17/21, 13/21, 0/21, 0/21, 0/21 positive pools for rmlB, rmlD, rmlA, rmlC and glf respectively (Table 4). For the wzx, wzy and three transferase genes there were no positives amongst the 21 pools of E. coli chromosomal DNA tested (Table 4). In each case the #44 pool gave a positive result.

B. Primers based on the E. coli 0111 O antigen gene cluster sequence.

One to four pairs of primers for each of the transferase, *wzx* and *wzy* genes of 0111 were tested against 5 the pools 1 to 21 of E. coli chromosomal DNA (Table 5). For *wbdH*, four pairs of primers, which bind to various regions of this gene, were tested and found to be specific for 0111 as there was no amplified DNA of the correct size in any of those 21 pools of E. coli chromosomal DNA. 10 tested. Three pairs of primers for *wbdM* were tested, and they are all specific although primers #985/#986 produced a band of the wrong size from one pool. Three pairs of primers for *wzx* were tested and they all were specific. Two pairs of primers were tested for *wzy*, both are 15 specific although #980/#983 gave a band of the wrong size in all pools. One pair of primers for *wbdL* was tested and found unspecific and therefore no further test was carried out. Thus, *wzx*, *wzy* and two of the three transferase genes are highly specific to 0111. Bands of the wrong 20 size found in amplified DNA are assumed to be due to chance hybridisation of genes widely present in E. coli. The primers, annealing temperatures and positions for each gene are in (Table 5).

The 0111 assay was also performed using pools 25 including DNA from O antigen expressing Yersinia pseudotuberculosis, Shigella boydii and Salmonella enterica strains (Table 5A). None of the oligonucleotides derived from *wbdH*, *wzx*, *wzy* or *wbdM* gave amplified DNA of the correct size with these pools. Notably, pool number 30 25 includes S. enterica Adelaide which has the same O antigen as E. coli 0111: this pool did not give a positive PCR result for any primers tested indicating that these genes are highly specific for E. coli 0111.

Each of the 12 pairs binding to *wbdH*, *wzx*, *wzy* and 35 *wbdM* produces a band of predicted size with the pools containing 0111 DNA (pools number 22 to 24). As pools 22 to 24 included DNA from all strains present in pool 21 plus 0111 strain DNA (Table 2), we conclude that the 12

pairs of primers all give a positive PCR test with each of three unrelated 0111 strains but not with any other strains tested. Thus these genes are highly specific for E. coli 0111.

5

C. Primers based on the E. coli 0157 O antigen gene cluster sequence.

Two or three primer pairs for each of the transferase, wzx and wzy genes of 0157 were tested against 10 E. coli chromosomal DNA of pools 1 to 19, 29 and 30 (Table 6). For *wbdN*, three pairs of primers, which bind to various regions of this gene, were tested and found to be specific for 0157 as there was no amplified DNA in any of those 21 pools of E. coli chromosomal DNA tested. Three 15 pairs of primers for *wbdO* were tested, and they are all specific although primers # 1211/#1212 produced two or three bands of the wrong size from all pools. Three pairs of primers were tested for *wbdP* and they all were specific. Two pairs of primers were tested for *wbdR* and 20 they were all specific. For *wzy*, three pairs of primers were tested and all were specific although primer pair #1203/#1204 produced one or three bands of the wrong size in each pool. For *wzx*, two pairs of primers were tested and both were specific although primer pair #1217/#1218 25 produced 2 bands of wrong size in 2 pools, and 1 band of wrong size in 7 pools. Bands of the wrong size found in amplified DNA are assumed to be due to chance hybridisation of genes widely present in E. coli. The primers, annealing temperatures and function information 30 for each gene are in Table 6.

The 0157 assay was also performed using pools 37 to 42, including DNA from O antigen expressing Yersinia pseudotuberculosis, Shigella boydii, Yersinia enterocolitica 09, Brucella abortus and Salmonella enterica strains (Table 6A). None of the oligonucleotides 35 derived from *wbdN*, *wzy*, *wbdO*, *wzx*, *wbdP* or *wbdR* reacted specifically with these pools, except that primer pair #1203/#1204 produced two bands with Y. enterocolitica 09

and one of the bands is of the same size with that from the positive control. Primer pair #1203/#1204 binds to wzy. The predicted secondary structures of Wzy proteins are generally similar, although there is very low 5 similarity at amino acid or DNA level among the sequenced wzy genes. Thus, it is possible that Y. enterocolitica 09 has a wzy gene closely related to that of E. coli 0157. It is also possible that this band is due to chance 10 hybridization of another gene, as the other two wzy primer pairs (#1205/#1206 and #1207/#1208) did not produce any band with Y. enterocolitica 09. Notably, pool number 37 includes S. enterica Landau which has the same O antigen 15 includes S. enterica Landau which has the same O antigen as E. coli 0157, and pool 38 and 39 contain DNA of B. abortus and Y. enterocolitica 09 which cross react serologically with E. coli 0157. This result indicates that these genes are highly 0157 specific, although one primer pair may have cross reacted with Y. enterocolitica 09.

Each of the 16 pairs binding to *wbdN*, *wzx*, *wzy*, *wbdO*, 20 *wbdP* and *wbdR* produces a band of predicted size with the pools containing 0157 DNA (pools number 31 to 36). As pool 29 included DNA from all strains present in pools 31 to 36 other than 0157 strain DNA (Table 3), we conclude 25 that the 16 pairs of primers all give a positive PCR test with each of the five unrelated 0157 strains.

Thus PCR using primers based on genes *wbdN*, *wzy*, *wbdO*, *wzx*, *wbdP* and *wbdR* is highly specific for E. coli 0157, giving positive results with each of six unrelated 30 0157 strains while only one primer pair gave a band of the expected size with one of three strains with O antigens known to cross-react serologically with E. coli 0157.

D. Primers based on the Salmonella enterica serotype C2 and B O antigen gene cluster sequences.

35 We also performed a PCR using primers for the S. enterica C2 and B serogroup transferases, *wzx*, *wzy* and genes (Tables 7 to 9). The nucleotide sequences of C2

and B O antigen gene clusters are listed as SEQ ID NO: 3 (Fig. 9) and SEQ ID NO:4 (Fig. 10) respectively.

5 Chromosomal DNA from all the 46 serotypes of Salmonella enterica (Table 9) was isolated using the Promega Genomic isolation kit, 7 pools of 4 to 8 samples per pool were made. Salmonella enterica serotype B or C2 DNA was omitted from the pool for testing primers of 46 respective serotypes but added to a pool containing 6 other samples to give pool number 8 for use as a positive control.

10 PCR reactions were carried out under the following conditions: denaturing, 94°C/30"; annealing, temperature varies (see below)/30"; extension, 72°C/1'; 30 cycles. PCR reaction was carried out in a volume of 25µL for each pool. After the PCR reaction, 10µL PCR product from each 15 pool was run on an agarose gel to check for amplified DNA. For pools which gave a band of correct size, PCR was repeated using individual chromosomal samples of that pool, and agarose gel was run to check for amplified DNA from each sample.

20 The Salmonella enterica serotype B O antigen gene cluster (of strain LT2) was the first O antigen gene cluster to be fully sequenced, and the function of each gene has been identified experimentally [Jiang, X. M., Neal, B., Santiago, F., Lee, S. J., Romana, L. K., and Reeves, P. R. (1991) "Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar *typhimurium* (strain LT2)." *Mol. Microbiol.* **5**(3), 695-713; Liu, D., Cole, R., and Reeves, P. R. (1996). "An O antigen processing function for Wzx(RfbX): a promising candidate 25 for O-unit flippase" *J. Bacteriol.*, **178**(7), 2102-2107; Liu, D., Haase, A. M., Lindqvist, L., Lindberg, A. A., and Reeves, P. R. (1993). "Glycosyl transferases of O-antigen 30 biosynthesis in *S. enterica* : identification and characterisation of transferase genes of groups B, C2 and E1." *J. Bacteriol.*, **175**, 3408-3413; Liu, D., Lindquist, L., and Reeves P. R. (1995). "Transferases of O-antigen 35 biosynthesis in *Salmonella enterica*: dideoxhexosyl

transferases of groups B and C2 and acetyltransferase of group C2." J. Bacteriol., 177, 4084-4088; Romana, L. K., Santiago, F. S., and Reeves, P. R. (1991). "High level expression and purification dThymidine-diphospho-D-glucose 5 4,6 dehydratase (*rfbB*) from *Salmonella* serovar *typhimurium* LT2." BBRC, 174, 846-852]. One pair of primers for each of the pathway genes and *wbaP* was tested against the pools of *Salmonella enterica* DNA, two to three pairs of primers for each of the other transferases and *wzx* genes were also 10 tested. See Table 8 for a list of primers and functional information of each gene, as well as the annealing temperature of the PCR reaction for each pair of primers.

For pathway genes of group B strain LT2, there are 19/45, 14/45, 15/45, 12/45, 6/45, 6/45, 6/45, 6/45, 1/45, 15 9/45, 8/45 positives for *rmlB*, *rmlD*, *rmlA*, *rmlC*, *ddhD*, *ddhA*, *ddhB*, *ddhC*, *abe*, *manC*, and *manB* respectively (Table 9).

For the LT2 *wzx* gene we used three primer pairs each of which gave 1/45 positive. For the 4 transferase genes 20 we used a total of 9 primer pairs. 2 primer pairs for *wbaV* gave 2/90 positives. For 3 primer pairs of *wbaN*, 11/135 gave a positive result. For the *wbaP* primer pair 10/45 gave a positive result (Table 9).

The experimental data show that oligonucleotides 25 derived from the *wzx* and *wbaV* group B O antigen genes are specific for group B O antigen amongst all 45 *Salmonella enterica* O antigen groups except O group 67. The oligonucleotides derived from *Salmonella enterica* B group *wbaN* and *wbaU* genes detected B group O antigen and also 30 produced positive results with groups A, D1 and D3. *WbaU* encodes a transferase for a Mannose $\alpha(1-4)$ Mannose linkage and is expressed in groups A, B and D1 while *wbaN*, which encodes a transferase for Rhamnose $\alpha(1-3)$ Galactose 35 linkage is present in groups A, B, D1, D2, D3 and E1. This accounts for the positive results with the group B *wbaU* and *wbaN* genes. The *wbaN* gene of groups E and D2 has considerable sequence differences from that of groups A,

B, D1 and D3 and this accounts for the positive results only with groups B, D1 and D3.

The Salmonella enterica B primers derived from wzx and transferase genes produced a positive result with Salmonella enterica 067. We find that Salmonella enterica 067 has all the genes of the group B O antigen cluster.

There are several possible explanations for this finding including the possibility that the gene cluster is not functional due to mutation and the group 067 antigenicity is due to another antigen, or the O antigen is modified after synthesis such that its antigenicity is changed.

Salmonella enterica 067 would therefore be scored as Salmonella enterica group B in the PCR diagnostic assay.

However, this is of little importance because Salmonella enterica 067 is a rare O antigen and only one (serovar Crossness) of the 2324 known serovars has the 067 serotype [Popoff M.Y. et al (1992) "Antigenic formulas of the Salmonella enterica serovars" 6th revision WHO Collaborating Centre for Reference and Research on Salmonella enterica, Institut Pasteur Paris France], and serovar Crossness had only been isolated once [M. Popoff, personal communication].

The Salmonella enterica B primers derived from wbaP reacted with group A, C2, D1, D2, D3, E1, 54, 55, 67 and E4 O antigen groups. WbaP encodes the galactosyl transferase which initiates O unit synthesis by transfer of Galactose phosphate to the lipid carrier Undecaprenol phosphate. This reaction is common to the synthesis of several O antigens. As such wbaP is distinguished from other transferases of the invention as it does not make a linkage within an O antigen.

We also tested 20 primer pairs for the wzx, wzy and 5 transferase genes of serotype C2 and found no positives in all the 7 pools (Table 7).

Groups A, B, D1, D2, D3, C2 and E1 share many genes in common. Some of these genes occur with more than one sequence in which case each specific sequence can be named after one of the serogroups in which it occurs. The

distribution of these sequence specificities is shown in Table 10. The inventors have aligned the nucleotide sequences of Salmonella enterica wzy, wzx genes and transferase genes so as to determine specific combinations of nucleic acid molecules which can be employed to specifically detect and identify the Salmonella enterica groups A, B, D1, D2, D3, C2 and E1 (Table 10). The results show that many of the O antigen groups can be detected and identified using a single specific nucleic acid molecule although other groups in particular D2 and E1, and A and D1 require a panel of nucleic acid molecules derived from a combination of genes.

It will be understood that in carrying out the methods of the invention with respect to the testing of particular sample types including samples from food, patients and faeces the samples are prepared by routine techniques routinely used in the preparation of such samples for DNA based testing.

TABLE 1

Pool No.	Strains of which chromosomal DNA included in the pool	Source*
1	<i>E. coli</i> type strains for O serotypes 1, 2, 3, 4, 10, 16, 18 and 39	IMVS ^a
2	<i>E. coli</i> type strains for O serotypes 40, 41, 48, 49, 71, 73, 88 and 100	IMVS
3	<i>E. coli</i> type strains for O serotypes 102, 109, 119, 120, 121, 125, 126 and 137	IMVS
4	<i>E. coli</i> type strains for O serotypes 138, 139, 149, 7, 5, 6, 11 and 12	IMVS
5	<i>E. coli</i> type strains for O serotypes 13, 14, 15, 17, 19ab, 20, 21 and 22	IMVS
6	<i>E. coli</i> type strains for O serotypes 23, 24, 25, 26, 27, 28, 29 and 30	IMVS
7	<i>E. coli</i> type strains for O serotypes 32, 33, 34, 35, 36, 37, 38 and 42	IMVS
8	<i>E. coli</i> type strains for O serotypes 43, 44, 45, 46, 50, 51, 52 and 53	IMVS
9	<i>E. coli</i> type strains for O serotypes 54, 55, 56, 57, 58, 59, 60 and 61	IMVS
10	<i>E. coli</i> type strains for O serotypes 62, 63, 64, 65, 66, 68, 69 and 70	IMVS
11	<i>E. coli</i> type strains for O serotypes 74, 75, 76, 77, 78, 79, 80 and 81	IMVS
12	<i>E. coli</i> type strains for O serotypes 82, 83, 84, 85, 86, 87, 89 and 90	IMVS
13	<i>E. coli</i> type strains for O serotypes 91, 92, 95, 96, 97, 98, 99 and 101	IMVS
14	<i>E. coli</i> type strains for O serotypes 103, 104, 105, 106, 107, 108 and 110	IMVS
15	<i>E. coli</i> type strains for O serotypes 112, 162, 113, 114, 115, 116, 117 and 118	IMVS
16	<i>E. coli</i> type strains for O serotypes 123, 165, 166, 167, 168, 169, 170 and 171	See b
17	<i>E. coli</i> type strains for O serotypes 172, 173, 127, 128, 129, 130, 131 and 132	See c
18	<i>E. coli</i> type strains for O serotypes 133, 134, 135, 136, 140, 141, 142 and 143	IMVS
19	<i>E. coli</i> type strains for O serotypes 144, 145, 146, 147, 148, 150, 151 and 152	IMVS

*

- a. Institute of Medical and Veterinary Science, Adelaide, Australia
- b. 123 from IMVS; the rest from Statens Serum Institut, Copenhagen, Denmark
- c. 172 and 173 from Statens Serum Institut, Copenhagen, Denmark, the rest from IMVS

- 50 -

TABLE 2

Pool No.	Strains of which chromosomal DNA included in the pool	Source*
20	<i>E. coli</i> type strains for O serotypes 153, 154, 155, 156, 157, 158, 159 and 160	IMVS
21	<i>E. coli</i> type strains for O serotypes 161, 163, 164, 8, 9 and 124	IMVS
22	As pool #21, plus <i>E. coli</i> 0111 type strain Stoke W.	IMVS
23	As pool #21, plus <i>E. coli</i> 0111:H2 strain C1250-1991	See d
24	As pool #21, plus <i>E. coli</i> 0111:H12 strain C156-1989	See e
25	As pool #21, plus <i>S. enterica</i> serovar Adelaide	See f
26	<i>Y. pseudotuberculosis</i> strains of O groups IA, IIA, IIB, IIC, III, IVA, IVB, VA, VB, VI and VII	See g
27	<i>S. boydii</i> strains of serogroups 1, 3, 4, 5, 6, 8, 9, 10, 11, 12, 14 and 15	See h
28	<i>S. enterica</i> strains of serovars (each representing a different O group) Typhi, Montevideo, Ferruch, Jangwani, Raus, Hvittingfoss, Waycross, Dan, Dugbe, Basel, 65,:i:e,n,z,15 and 52:d:e,n,x,z15	IMVS

*

- d. C1250-1991 from Statens Serum Institut, Copenhagen, Denmark
- e. C156-1989 from Statens Serum Institut, Copenhagen, Denmark
- f. *S. enterica* serovar Adelaide from IMVS
- g. Dr S Aleksic of Institute of Hygiene, Germany
- h. Dr J Lefebvre of Bacterial Identification Section, Laboratoire de Santé Publique du Québec, Canada

TABLE 3

Pool No.	Strains of which chromosomal DNA included in the pool	Source*
29	<i>E. coli</i> type strains for O serotypes 153, 154, 155, 156, 158, 159 and 160	IMVS
30	<i>E. coli</i> type strains for O serotypes 161, 163, 164, 8, 9, 111 and 124	IMVS
31	As pool #29, plus <i>E. coli</i> O157 type strain A2 (O157:H19)	IMVS
32	As pool #29, plus <i>E. coli</i> O157:H16 strain C475-89	See d
33	As pool #29, plus <i>E. coli</i> O157:H45 strain C727-89	See d
34	As pool #29, plus <i>E. coli</i> O157:H2 strain C252-94	See d
35	As pool #29, plus <i>E. coli</i> O157:H39 strain C258-94	See d
36	As pool #29, plus <i>E. coli</i> O157:H26	See e
37	As pool #29, plus <i>S. enterica</i> serovar Landau	See f
38	As pool #29, plus <i>Brucella abortus</i>	See g See h
39	As pool #29, plus <i>Y. enterocolitica</i> O9	
40	<i>Y. pseudotuberculosis</i> strains of O groups IA, IIA, IIB, IIC, III, IVA, IVB, VA, VB, VI and VII	See i
41	<i>S. boydii</i> strains of serogroups 1, 3, 4, 5, 6, 8, 9, 10, 11, 12, 14 and 15	See j
42	<i>S. enterica</i> strains of serovars (each representing a different O group) Typhi, Montevideo, Ferruch, Jangwani, Raus, Hvittingfoss, Waycross, Dan, Dugbe, Basel, 65:i:e,n,z15 and 52:d:e,n,x,z15	IMVS
43	<i>E. coli</i> type strains for O serotypes 1,2,3,4,10,18 and 29	IMVS
44	As pool #43, plus <i>E. coli</i> K-12 strains C600 and WG1	IVMS See k

*

- d. O157 strains from Statens Serum Institut, Copenhagen, Denmark
- e. O157:H26 from Dr R Brown of Royal Children's Hospital, Melbourne, Victoria
- f. *S. enterica* serovar Landau from Dr M Poppoff of Institut Pasteur, Paris, France
- g. *B. Abortus* from the culture collection of The University of Sydney, Sydney, Australia
- h. *Y. enterocolitica* O9 from Dr. K. Bettelheim of Victorian Infectious Diseases Reference Laboratory Victoria, Australia.
- i. Dr S Aleksic of Institute of Hygiene, Germany
- j. Dr J Lefebvre of Bacterial Identification Section, Laboratoire de Santé Publique du Québec, Canada
- k. Strains C600 and WG1 from Dr. B.J. Backmann of Department of Biology, Yale University, USA.

TABLE 4 PCR assay result using primers based on the E. coli serotype O16 (strain K-12) O antigen gene cluster sequence

Gene	Function	Base positions of the gene	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (out of 21) giving band of correct size	Annealing temperature of the PCR
<i>rmlB</i> *	TDP-rihamnose pathway	90-1175	#1064(91-109)	#1065(1175-1157)	1085bp	17	60°C
<i>rmlD</i> *	TDP-rihamnose pathway	1175-2074	#1066(1175-1193)	#1067 (2075-2058)	901bp	13	60°C
<i>rmlA</i> *	TDP-rihamnose pathway	2132-3013	#1068(2131-2148)	#1069(3013-2995)	883bp	0	60°C
<i>rmlC</i> *	TDP-rihamnose pathway	3013-3570	#1070(3012-3029)	#1071(3570-3551)	559bp	0	60°C
<i>gff</i> *	Galactofuranose pathway	4822-5925	#1074(4822-4840)	#1075(5925-5908)	1104bp	0	55°C
<i>wzx</i> *	Flippase	3567-4814	#1072(3567-3586)	#1073(4814-4797)	1248bp	0	55°C
<i>wzy</i> *	O polymerase	5925-7091	#1076(5925-5944)	#1077(7091-7074)	1167bp	0	60°C
<i>wbbI</i> *	Galactofuranosyl transferase	7094-8086	#1078 (7094-7111)	#1079(8086-8069)	993bp	0	50°C
<i>wbbJ</i> *	Acetyltransferase	8067-8654	#1080(8067-8084)	#1081(8654-8632)	588bp	0	60°C
<i>wbbK</i> **	Glucosyl transferase	5770-6888	#1082(5770-5787)	#1083(6888-6871)	1119bp	0	55°C
<i>wbbL</i> ***	Rhamnosyltransferase	679-1437	#1084(679-697)	#1085(1473-1456)	795bp	0***	55°C

* , **, *** Base positions based on GenBank entry U09876, U03041 and L19537 respectively
 *** 19 pools giving a band of wrong size

TABLE 5 PCR assay data using 0111 primers

Gene	Base positions of the gene according to SEQ ID NO: 1	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (out of 21) giving band of correct size	Annealing temperature of the PCR
<i>wbdH</i>	739-1932	#866 (739-757)	#867(1941-1924)	1203bp	0	60°C
		#976(925-942)	#978(1731-1714)	807bp	0	60°C
		#976(925-942)	#979(1347-1330)	423bp	0	60°C
		#977(1165-1182)	#978(1731-1714)	567bp	0	60°C
<i>wzx</i>	8646-9911	#969(8646-8663)	#970(9908-9891)	1263bp	0	50°C
		#1060(8906-8923)	#1062(9468-9451)	563bp	0	60°C
		#1061(9150-9167)	#1063 (9754-9737)	605bp	0	50°C
<i>wzy</i>	9901-10953	#900(976-9996)	#901(10827-10807)	852bp	0	60°C
		#980(10113-10130)	#983(10484-10467)	372bp	0*	61°C
		#870(10931-10949)	#871(11824-11796)	894bp	7	60°C
<i>wbdL</i>	10931-11824	#870(10931-10949)	#983(10484-10467)	372bp	0*	61°C
<i>wbdM</i>	11821-12945	#868(11821-11844)	#869(12945-12924)	1125bp	0	60°C
		#984(12042-12059)	#987(12447-12430)	406bp	0	60°C
		#985(12258-12275)	#986(12698-12681)	441bp	0**	65°C

* Giving a band of wrong size in all pools
 ** One pool giving a band of wrong size

TABLE 5A PCR specificity test data using 0111 primers

Gene	Base positions of the gene according to SEQ ID NO: 1	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (pools no. 25-28) giving band of correct size	Annealing temperature of the PCR
<i>wbdH</i>	739-1932	#866 (739-757)	#867 (1941-1924)	1203bp	0*	60°C
		#976 (925-942)	#978 (1731-1714)	807bp	0	60°C
		#976 (925-942)	#979 (1347-1330)	423bp	0	60°C
		#977 (1165-1182)	#978 (1731-1714)	567bp	0	60°C
		#969 (8646-8663)	#970 (9908-9891)	1263bp	0	55°C
<i>w2x</i>	8646-9911	#1060 (8906-8923)	#1062 (9468-9451)	563bp	0	60°C
		#1061 (9150-9167)	#1063 (9754-9737)	605bp	0*	50°C
	9901-10953	#900 (9976-9996)	#901 (10827-10807)	852bp	0	60°C
<i>wzy</i>		#980 (10113-10130)	#983 (10484-10467)	372bp	0**	60°C
	10931-11824	#870 (10931-10949)	#871 (11824-11796)	894bp	0	60°C
<i>wbdL</i>	11821-12945	#868 (11821-11844)	#869 (12945-12924)	1125bp	0	60°C
<i>wbdM</i>		#984 (12042-12059)	#987 (12447-12430)	406bp	0	60°C
		#985 (12258-12275)	#986 (12698-12681)	441bp	0*	65°C

* 1 pool giving a band of wrong size

** 2 pools giving 3 bands of wrong sizes, 1 pool giving 2 bands of wrong sizes

TABLE 6 PCR results using primers based on the *E. coli* O157 sequence

Gene	Function	Base position of the gene according to SEQ ID NO: 2	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (out of 21) giving band of correct size	Annealing temperature of the PCR
<i>wbdN</i>	Sugar transferase	79-861	#1197(79-96)	#1198 (861-844)	783	0	55°C
			#1199(184-201)	#1200(531-514)	348	0	55°C
			#1201(310-327)	#1202(768-751)	459	0	55°C
<i>wzy</i>	O antigen	858-2042	#1203(838-875)	#1204(2042-2025)	1185	0*	50°C
			#1205(1053-1070)	#1206(1619-1602)	567	0	63°C
			#1207(1278-1295)	#1208(1913-1896)	636	0	60°C
			#1209(2011-2028)	#1210(2757-2740)	747	0	50°C
<i>wbdO</i>	Sugar transferase	2011-2757	#1211(2110-2127)	#1212(2493-2476)	384	0**	62°C
			#1213(2305-2322)	#1214(2682-2665)	378	0	60°C
			#1216(4135-4118)	1392	0	50°C	
<i>wzx</i>	O antigen flippase	2744-4135	#1215(2744-2761)	#1216(2493-2476)	687	0***	63°C
			#1217(2942-2959)	#1218(3628-3611)			
<i>wbdP</i>	Sugar transferase	5257-6471	#1221(5257-5274)	#1222(6471-6454)	1215	0	55°C
			#1223(5440-5457)	#1224(5973-5956)	534	0	55°C
			#1225(5707-5724)	#1226(6231-6214)	525	0	55°C
<i>wbdR</i>	N-acetyl transferase	13156-13821	#1229(13261-13278)	#1230(13629-13612)	369	0	55°C
			#1231(13384-13401)	#1232(13731-13714)	348	0	60°C

* 3 bands of wrong size in one pool, 1 band of wrong size in all other pools

** 3 bands of wrong sizes in 9 pools, 2 bands of wrong size in all other pools

*** 2 bands of wrong sizes in 2 pools, 1 band of wrong size in 7 pools

TABLE 6A PCR results using primers based on the *E. coli* O157 sequence

Gene	Function	Base position of the gene according to SEQ ID NO: 2	Forward primer (base positions)	Reverse primer (base positions)	Length of the PCR fragment	Number of pools (pools no. 37-42) giving band of correct size	Annealing temperature of the PCR
<i>wbdN</i>	Sugar transferase	79-861	#1197(79-96)	#1198 (861-844)	783	0*	55°C
			#1199(184-201)	#1200(531-514)	348	0*	55°C
			#1201(310-327)	#1202(768-751)	459	0	61°C
<i>wzy</i>	O antigen polymerase	858-2042	#1203(838-875)	#1204(2042-2025)	1185	1**	50°C
			#1205(1053-1070)	#1206(1619-1602)	567	0***	60°C
			#1207(1278-1295)	#1208(1913-1896)	636	0	60°C
<i>wbdO</i>	Sugar transferase	2011-2757	#1209(2011-2028)	#1210(2757-2740)	747	0	50°C
			#1211(2110-2127)	#1212(2493-2476)	384	0****	61°C
			#1213(2305-2322)	#1214(2682-2665)	378	0	60°C
<i>wzx</i>	O antigen flippase	2744-4135	#1215(2744-2761)	#1216(4135-4118)	1392	0	50°C
			#1217(2942-2959)	#1218(3628-3611)	687	0	63°C
<i>wbdP</i>	Sugar transferase	5257-6471	#1221(5257-5274)	#1222(6471-6454)	1215	0	55°C
			#1223(5440-5457)	#1224(5973-5956)	534	0*	60°C
			#1225(5707-5724)	#1226(6231-6214)	525	0	55°C
<i>wbdR</i>	N-acetyl transferase	13156-13821	#1229(13261-13278)	#1230(13629-	369	0	50°C
			#1231(13384-13401)	#1232(13731-	348	0	60°C

* 1 band of wrong size in one pool
 ** pool #39 giving two bands, one band of correct size, the other band of wrong size in another pool.

*** 2 bands of wrong sizes in one pool
 **** 3 bands of wrong sizes in 2 pools, 2 bands of wrong sizes in 2 other pools

TABLE 7
PCR assay data using primers based on the *Salmonella enterica* serotype C2 (strain M67)
O antigen gene cluster sequence

Gene	Function	Base positions of the gene according to SEQ ID NO: 3	Forward primer (base position)	Reverse primer (base position)	Length of the PCR fragment	Number of pools (out of 7) giving band of correct size	Annealing temperature of the PCR
wzx	Flippase	1019-2359	#1144(1019-1036)	#1145(1414-1397)	396bp	0	55°C
			#1146(1708-1725)	#1147(2170-2153)	463bp	0	55°C
			#1148(1938-1955)	#1149(2356-2339)	419bp	0	55°C
wbaR	Abequosyl transferase	2352-3314	#1150(2352-2369)	#1151(2759-2742)	408bp	0	55°C
			#1152(2601-2618)	#1153(3047-3030)	447bp	0	55°C
			#1154(2910-2927)	#1155(3311-3294)	402bp	0	55°C
wbaL	Acetyl transferase	3361-3875	#1156(3361-3378)	#1157(3759-3742)	399bp	0	55°C
			#1158(3578-3595)	#1159(3972-3955)	395bp	0	50°C
wbaQ	Rhamnosyl	3977-5020	#1160(3977-3994)	#1161(4378-4361)	402bp	0	55°C
			#1162(4167-4184)	#1163(4774-4757)	608bp	0	55°C
wzy	O polymerase	5114-6313	#1164(4603-4620)	#1165(5017-5000)	415bp	0*	60°C
			#1166(5114-5131)	#1167(5515-5498)	402bp	0**	55°C
wbaW	Mannosyl transferase	6313-7323	#1168(5664-5681)	#1169(6112-6095)	449bp	0	55°C
			#1170(5907-5924)	#1171(6310-6293)	404bp	0	55°C
			#1172(6313-6330)	#1173(6805-6788)	493bp	0	50°C
wbaZ	Mannosyl transferase	7310-8467	#1174(6697-6714)	#1175(7068-7051)	372bp	0	55°C
			#1176(6905-6922)	#1177(7320-7303)	416bp	0	55°C
			#1178(7310-7327)	#1179(7775-7758)	466bp	0	50°C
			#1180(7530-7547)	#1181(7907-7890)	378bp	0	55°C
			#1182(8007-8024)	#1183(8464-8447)	458bp	0	55°C

* Positive pool gives another band, which is also present in another pool. All other pools gave bands of wrong size.

** Band of wrong size in 6 other pools.

TABLE 8
PCR primers based on the *Salmonella enterica* serotype B (strain LT2) O antigen gene cluster sequence

Gene	Function	Base position of the gene according to SEQ ID NO: 4	Forward primer (base position)	Reverse primer (base position)	Length of the PCR fragment	Annealing temperature of the PCR
<i>rmlB</i>	TDP-rihamnose pathway	4099-5184	#1094 (4100-4117)	#1095(4499-4482)	400bp	55°C
<i>rmlD</i>	TDP-rihamnose pathway	5184-6083	#1092(5186-5203)	#1093(5543-5526)	358bp	50°C
<i>rmlA</i>	TDP-rihamnose pathway	6131-7009	#1090(6531-6548)	#1091(6837-6820)	308bp	55°C
<i>rmlC</i>	TDP-rihamnose pathway	7010-7561	#1088(7013-7030)	#1089(7372-7355)	360bp	55°C
<i>ddhD</i>	CDP-abequose pathway	7567-8559	#1112(7567-7584)	#1113(7970-7953)	404bp	55°C
<i>ddhA</i>	CDP-adequose pathway	8556-9329	#1114(8556-8573)	#1115(8975-8958)	420bp	60°C
<i>ddhB</i>	CDP-adequose pathway	9334-10413	#1116(9334-9351)	#1117(9816-9799)	483bp	45°C
<i>ddhC</i>	CDP-adequose pathway	10440-11753	#1118(10440-10457)	#1119(10871-10854)	432bp	60°C
<i>abe</i>	CDP-adequose pathway	11781-12680	#1100(12008-12025)	#1101(12388-12371)	381bp	55°C
<i>wzx</i>	Flippase	12762-14054	#1120(12762-12779)	#1121(13150-13133)	389bp	55°C
<i>wbaV</i>	Abequosyl transferase	14059-15060	#1122(12993-13010)	#1123(13417-13400)	425bp	55°C
<i>wbaU</i>	Mannosyl transferase	15379-16440	#1124(13635-13652)	#1125(14051-14034)	417bp	55°C
<i>wbaN</i>	Rhamnosyl transferase	16441-17385	#1126(14059-14076)	#1127(14421-14404)	363bp	45°C
			#1128(14688-14705)	#1129(15057-15040)	370bp	45°C
			#1130(15379-15396)	#1131(15768-15751)	390bp	60°C
			#1132(15850-15867)	#1133(16262-16245)	413bp	50°C
			#1134(16027-16044)	#1135(16437-16420)	411bp	60°C
			#1136(16441-16458)	#1137(16851-16834)	411bp	45°C
			#1138(16630-16647)	#1139(17087-17070)	458bp	55°C
			#1140(16978-16995)	#1141(17382-17365)	405bp	50°C
<i>manC</i>	GDP-mannose pathway	17386-18825	#1098(17457-17474)	#1099(18143-18126)	687bp	60°C
<i>mabB</i>	GDP-mannose pathway	18812-20245	#1096(18991-19008)	#1097(19345-19328)	355bp	55°C
<i>wbaP</i>	Galactosyl transferase	20317-21747	#1142(20389-20406)	#1143(20709-20692)	321bp	55°C

TABLE 9 PCR results using LT2 primers*

* y indicates a positive PCR result. Blank indicates a negative result.

TABLE 10 Gene specificities in *Salmonella enterica* serogroups

Serogroup	Genes										
	wzy	wzx	wbaP	wbaU	wbaN	wbaV	wbaO	wbaW	wbaZ	wbaQ	wbaR
A	B	D	B	B	B	D	-	-	-	-	-
B	B	B	B	B	B	B	-	-	-	-	-
D1	B	D	B	B	B	D	-	-	-	-	-
D2	E1	D	B	-	E1	D	E1	-	-	-	-
D3	D3	D	B	B	B	D	-	-	-	-	-
C2	C2	C2	B	-	-	-	C2	C2	C2	C2	C2
E1	E1	E1	B	-	E1	-	E1	-	-	-	-

- means 'not present'

- 61 -

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Reeves, Peter R
Wang, Lei
- (ii) TITLE OF INVENTION: Nucleic Acid Molecules Specific For
Bacterial Antigens And Uses Thereof
- (iii) NUMBER OF SEQUENCES: 4
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Thomas Gumley
 - (B) STREET: 168 Walker Street
 - (C) CITY: North Sydney
 - (D) STATE: New South Wales
 - (E) COUNTRY: Australia
 - (F) ZIP: 2068
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Gumley, Thomas P
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 99575944
 - (B) TELEFAX: 99576288

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 14516 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: YES
- (v) ORIGINAL SOURCE:
 - (A) ORGANISM: Escherichia coli

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GATCTGATGG CCGTAGGGCG CTACGTGCTT TCTGCTGATA TCTGGGCTGA GTTGGAAAAA	60
ACTGCTCCAG GTGCCTGGGG ACGTATTCAA CTGACTGATG CTATTGCAGA GTTGGCTAAA	120
AAACAGTCTG TTGATGCCAT GCTGATGACC GGCGACAGCT ACGACTGCGG TAAGAAGATG	180

GGCTATATGC AGGCATTCTGT TAAGTATGGG CTGCGCAACC TTAAAGAAGG GGCGAAGTTC	240
CGTAAGAGCA TCAAGAAGCT ACTGAGTGAG TAGAGATTAA CACGTCTTG TGACGATAAG	300
CCAGAAAAAA TAGCGGCAGT TAACATCCAG GCTTCTATGC TTTAAGCAAT GGAATGTTAC	360
TGCCGTTTT TATGAAAAAT GACCAATAAT AACAAAGTTAA CCTACCAAGT TTAATCTGCT	420
TTTGTTGGA TTTTTCTTG TTTCTGGTCG CATTGGTAA GACAATTAGC GTGAGTTTA	480
GAGAGTTTG CGGGATCTCG CGGAAC TGCT CACATCTTG GCATTAGTT AGTGCAGTGG	540
TAGCTGTTAA GCCAGGGCG GTAGCTGCC TAATTAATT TTAACGTATA CATTATTCT	600
TGCCGTTAT AGCAAATAAA GTCAATCGGA TTAAACTTCT TTTCCATTAG GTAAAAGAGT	660
GTTTGTAGTC GCTCAGGGAA ATTGGTTTG GTAGTAGTAC TTTCAAATT ATCCATTTC	720
CGATTTAGAT GGCAGTTGAT GTTACTATGC TGCAACATA TCAATGTATA TTATTTACTT	780
TTAGAATGTG ATATGAAAAA AATAGTGATC ATAGGCAATG TAGCGTCAAT GATGTTAAGG	840
TTCAGGAAAG AATTAATCAT GAATTTAGTG AGGCAAGGTG ATAATGTATA TTGTCTAGCA	900
AATGATTTT CCACTGAAGA TCTTAAAGTA CTTTCGTCA GGGCGTTAA GGGGGTTAAA	960
TTCTCTCTTA ACTCAAAGGG TATTAATCCT TTTAAGGATA TAATTGCTGT TTATGAACTA	1020
AAAAAAATTC TTAAGGATAT TTCCCCAGAT ATTGTATTTT CATATTTGT AAAGCCAGTA	1080
ATATTTGGAA CTATTGCTTC AAAGTTGTCA AAAGTGCCTA GGATTGTTGG AATGATTGAA	1140
GGTCTAGGTA ATGCCTTCAC TTATTATAAG GGAAAGCAGA CCACAAAAAC TAAAATGATA	1200
AAGTGGATAC AAATTCTTT ATATAAGTTA GCATTACCGA TGCTTGATGA TTTGATTCTA	1260
TTAAATCATG ATGATAAAAAG ATGATAATC GATCAGTATA ATATTAAGC TAAGGTAACA	1320
GTGTTAGGTG GGATTGGATT GGATCTTAAT GAGTTTCAT ATAAAGAGCC ACCGAAAGAG	1380
AAAATTACCT TTATTTTAT AGCAAGGTTA TTAAGAGAGA AAGGGATATT TGAGTTATT	1440
GAAGCCCAA AGTTCGTTAA GACAACCTAT CCAAGTTCTG AATTGTAAT TTTAGGAGGT	1500
TTTGAGAGTA ATAATCCTTT CTCATTACAA AAAATGAAA TTGAATCGCT AAGAAAAGAA	1560
CATGATCTTA TTTATCCTGG TCATGTGGAA AATGTTCAAG ATTGTTAGA GAAAAGTTCT	1620
GTTTTGTTT TACCTACATC ATATCGAGAA GGCGTACCAA GGGTGATCCA AGAAGCTATG	1680
GCTATTGGTA GACCTGTAAT AACAACTAAT GTACCTGGGT GTAGGGATAT AATAAATGAT	1740
GGGGTCAATG GCTTTTGAT ACCTCCATT GAAATTAAATT TACTGGCAGA AAAATGAAA	1800
TATTTTATTG AGAATAAAGA TAAAGTACTC GAAATGGGC TTGCTGGAAG GAAGTTGCA	1860
GAAAAAAACT TTGATGCTTT TGAAAAAAAT AATAGACTAG CATCAATAAT AAAATCAAAT	1920
AATGATTTT GACTTGAGCA GAAATTATTT ATATTCAT CTGAAAATA AAGGCTGTTA	1980
TTATGAATAA AGTGGCATTAA ATTACTGGTA TCACTGGCA AGATGGCTCC TATTTGGCAG	2040
AATTATTGTT AGAAAAAGGT TATGAAGTTC ATGGTATTAA ACGCCGTGCA TCTTCATTAA	2100
ATACTGAGCG AGTGGATCAC ATCTATCAGG ATTACACATT AGCTAATCCT AAACTTTTC	2160
TACACTATGG CGATTTGACA GATACTTCCA ATCTGACCCG TATTTAAAAA GAAGTTCAAC	2220

- 63 -

CAGATGAAGT TTACAATTTG GGGGCGATGA	GCCATGTAGC GGTATCATT GAGTCACCA	2280
AATACACTGC TGATGTTGAT GCGATAGGAA	CATTGCGTCT TCTTGAAGCT ATCAGGATAT	2340
TGGGGCTGGA AAAAAAGACA	AAATTTATC AGGCTTCAAC TTCAGAGCTT TATGGTTGG	2400
TTCAAGAAAT TCCACAAAAA	GAGACTACGC CATTATC ACGTTCGCCT TATGCTGTTG	2460
CAAAATTATA TGCCTATTGG	ATCACTGTTA ATTATCGTGA GTCTTATGGT ATGTTTGCT	2520
GCAATGGTAT TCTCTTAAC	CACGAATCAC CTCGCCGTGG CGAGACCTTT GTTACTCGTA	2580
AAATAACACG CGGGATAGCA	AATATTGCTC AAGGTCTTGA TAAATGCTTA TACTTGGAA	2640
ATATGGATTC TCTGCGTGAT	TGGGGACATG CTAAGGATTA TGTCAAAATG CAATGGATGA	2700
TGCTGCAGCA AGAAAATCCA	GAAGATTTG TAATTGCTAC AGGAATTCAA TATTCTGTCC	2760
GTGAGTTGT CACAATGGCG	GCAGAGCAAG TAGGCATAGA GTTAGCATT GTAGGTGAGG	2820
GAGTAAATGA AAAAGGTGTT	GTTGTTCGG TCAATGGCAC TGATGCTAAA GCTGTAAACC	2880
CGGGCGATGT AATTATATCT	GTAGATCCAA GGTATTTAG GCCTGCAGAA GTTGAAACCT	2940
TGCTTGGCGA TCCTACTAAT	GCGCATAAAA AATTAGGATG GAGCCCTGAA ATTACATTGC	3000
GTGAAATGGT AAAAGAAATG	GTTTCCAGCG ATTTAGCAAT AGCGAAAAAG AACGTCTTGC	3060
TGAAAGCTAA TAACATTGCC	ACTAATATTC CGCAAGAATA AAAAGATAA TACATTAAAT	3120
AATTAAAAAT GGTGCTAGAT	TTATTAGTAC CATTATTTT TTTTGGGTGA CTAATGTTA	3180
TTACATCAGA TAAATTTAGA	GAAATTATCA AGTTAGTTCC ATTAGTATCA ATTGATCTGC	3240
TAATTGAAAA CGAGAATGGT	GAATATTTAT TTGGTCTTAG GAATAATCGA CGGGCCAAAA	3300
ATTATTTTT TGTCCAGGT	GGTAGGATTG CAAAAATGA ATCTATTAAA AATGCTTTA	3360
AAAGAATATC ATCTATGGAA	TTAGGTAAAG AGTATGGTAT TTCAGGAAGT GTTTTAATG	3420
GTGTATGGGA ACATTTCTAT	GATGATGGTT TTTTTCTGA AGGCGAGGCA ACACATTATA	3480
TAGTGCTTG TTACACACTG	AAAGTCTTA AAAGTGAATT GAATCTCCC GATGATCAAC	3540
ATCGTGAATA CCTTGGCTA	ACTAAACACC AAATAAATGC TAAACAAGAT GTTCATAACT	3600
ATTCAAAAAA TTATTTTTG	TAATTTTAT TAAAAATTAA TATGCGAGAG AATTGTATGT	3660
CTCAATGTCT TTACCTGTG	ATTATTGCCG GAGGAACCGG AAGCCGTCTA TGGCCGTTGT	3720
CTCGAGTATT ATACCCTAAA	CAATTTTAA ATTTAGTTGG GGATTCTACA ATGTTGAAA	3780
CAACAATTAC GCGTTGGAT	GGCATCGAAT GCGAAAATCC AATTGTTATC TGCAATGAAG	3840
ATCACCGATT TATTGTAGCA	GAGCAATTAC GACAGATTGG TAAGCTAACC AAGAATATTA	3900
TACTTGAGCC GAAAGGCCGT	AATACTGCAC CTGCCATAGC TTTAGCTGCT TTTATCGCTC	3960
AGAAGAATAA TCCTAATGAC	GACCCTTAT TATTAGTACT TGCGGCAGAC CACTCTATAA	4020
ATAATGAAAA AGCATTTCGA	GAGTCATAAA TAAAAGCTAT GCCGTATGCA ACTTCTGGGA	4080
AGTTAGTAAC ATTTGGAATT	ATTCCGGACA CGGCAAATAC TGGTTATGGA TATATTAAGA	4140
GAAGTTCTTC AGCTGATCCT	AATAAAGAAT TCCCAGCATA TAATGTTGCG GAGTTGTAG	4200
AAAAACCAGA TGTTAAAACA	GCACAGGAAT ATATTCGAG TGGGAATTAT TACTGGAATA	4260

- 64 -

GGCGAATGTT	TTTATTCGC	GCCAGTAAAT	ATCTTGATGA	ACTACGGAAA	TTTAGACCAG	4320
ATATTTATCA	TAGCTGTGAA	TGTGCAACCG	CTACAGCAAA	TATAGATATG	GACTTTGTCC	4380
GAATTAACGA	GGCTGAGTTT	ATTAATTGTC	CTGAAGAGTC	TATCGATTAT	GCTGTGATGG	4440
AAAAAAACAAA	AGACGCTGTA	GTTCTTCGA	TAGATATTGG	CTGGAATGAC	GTGGGTTCTT	4500
GGTCATCACT	TTGGGATATA	AGCCAAAAGG	ATTGCCATGG	TAATGTGTGC	CATGGGGATG	4560
TGCTCAATCA	TGATGGAGAA	AATAGTTTA	TTTACTCTGA	GTCAAGTCTG	GTGCGACAG	4620
TCGGAGTAAG	TAATTTAGTA	ATTGTCCAAA	CCAAGGATGC	TGTACTGGTT	GCGGACCGTG	4680
ATAAAGTCCA	AAATGTAAA	AACATAGTTG	ACGATCTAAA	AAAGAGAAAA	CGTGCTGAAT	4740
ACTACATGCA	TCGTGCAGTT	TTTCGCCCTT	GGGGTAAATT	CGATGCAATA	GACCAAGGCG	4800
ATAGATATAG	AGTAAAAAAA	ATAATAGTTA	AACCAGGAGA	AGGGTTAGAT	TTAAGGATGC	4860
ATCATCATAG	GGCAGAGCAT	TGGATTGTTG	TATCCGGTAC	TGCTAAAGTT	TCACTAGGTA	4920
GTGAAGTTAA	ACTATTAGTT	TCTAATGAGT	CTATATATAT	CCCTCAGGGA	GCAAAATATA	4980
GTCTTGAGAA	TCCAGGCGTA	ATACCTTG	ATCTAATTGA	AGTAAGTTCT	GGTGATTACC	5040
TTGAATCAGA	TGATATAGTG	CGTTTACTG	ACAGATATAA	CAGTAAACAA	TTCTAAAGC	5100
GAGATTGATA	AATATGAATA	AAATAACTTG	CTTCAAAGCA	TATGATATAC	GTGGCGTCT	5160
TGGTGCTGAA	TTGAATGATG	AAATAGCATA	TAGAATTGGT	CGCGCTTATG	GTGAGTTTT	5220
TAAACCTCAA	ACTGTAGTTG	TGGGAGGAGA	TGCTCGCTTA	ACAAGTGAGA	GTAAAGAGAA	5280
ATCACTCTCA	AATGGCTAT	GTGATGCAGG	CGTAAATGTC	TTAGATCTTG	GAATGTGTGG	5340
TACTGAAGAG	ATATATTTT	CCACTTGGTA	TTAGGAATT	GATGGTGGAA	TCGAGGTAAC	5400
TGCAAGCCAT	AATCCAATTG	ATTATAATGG	AATGAAATTA	GTAACCAAAG	GTGCTCGACC	5460
AATCAGCACT	GACACAGGTC	TCAAAGATAT	ACAACAATTA	GTAGAGAGTA	ATAATTGTA	5520
AGAGCTAAC	CTAGAAAAAA	AAGGGAATAT	TACCAAATAT	TCCACCCGAG	ATGCCTACAT	5580
AAATCATTG	ATGGGCTATG	CTAATCTGCA	AAAAATAAAA	AAAATCAAAA	TAGTTGTGAA	5640
TTCTGGGAAT	GGTGCAGCTG	GTCCTGTTAT	TGATGCTATT	GAGGAATGCT	TTTACGGAA	5700
CAATATCCG	ATTCAGTTG	AAAAAATAAA	TAATACACCC	GATGTAATT	TTCCACATGG	5760
TATCCCTAAT	CCATTACTAC	CTGAGTGCAG	AGAAGATACC	AGCAGTGCAG	TTATAAGACA	5820
TAGTGCTGAT	TTTGGTATTG	CATTGATGG	TGATTTGAT	AGGTGTTTT	TCTTGATGA	5880
AAATGGACAA	TTTATTGAAG	GATACTACAT	TGTTGGTTA	TTAGCGGAAG	TTTTTTAGG	5940
GAAATATCCA	AAAGCAAAAA	TCATTCAATGA	TCCTCGCCTT	ATATGGAATA	CTATTGATAT	6000
CGTAGAAAGT	CATGGTGGTA	TACCTATAAT	GAATGAAACC	GGTCATGCTT	ACATTAAGCA	6060
AAGAATGCGT	GAAGAGGATG	CCGTATATGG	CGGCGAAATG	AGTGCACATC	ATTATTTAA	6120
AGATTTGCA	TACTGCGATA	GTGGAATGAT	TCCTTGATT	TTAATTGAG	AACTTTGAG	6180
TCTGACAAAT	AAAAAATTAG	GTGAACCTGGT	TTGTGGTTGT	ATAAACGACT	GGCCGGCAAG	6240
TGGAGAAATA	AACTGTACAC	TAGACAATCC	GCAAAATGAA	ATAGATAAAAT	TATTTAATCG	6300

TTACAAAGAT AGTGCCTAG CTGTTGATTA CACTGATGGA TTAACATATGG AGTTCTCTGA	6360
TTGGCGTTT AATGTTAGAT GCTCAAATAC AGAACCTGTA GTACGATTGA ATGTAGAAC	6420
TAGGAATAAT GCTATTCTTA TGCAGGAAAA AACAGAAGAA ATTCTGAATT TTATATCAA	6480
ATAAAATTGC ACCTGAGTTC ATAATGGAA CAAGAAATAT ATGAAAGTAC TTCTGACTGG	6540
CTCAACTGGC ATGGTTGGTA AGAATATATT AGAGCATGAT AGTGAAGTA AATATAATAT	6600
ACTTACTCCA ACCAGCTCTG ATTTGAATT ATTAGATAAA AATGAAATAG AAAAATTCA	6660
GCTTATCAAC ATGCCAGACT GTATTATACA TGCAGCGGGG TTAGTGGAG GCATTATGC	6720
AAATATAAGC AGGCCGTTG ATTTCTGGA AAAAAATTG CAGATGGGTT TAAATTTAGT	6780
TTCCGTCGCA AAAAAACTAG GTATCAAGAA AGTGCTTAAC TTGGTAGTT CATGCATGTA	6840
CCCCAAAAAC TTTGAAGAGG CTATTCTGA GAAAGCTCTG TTAACCTGGAG AGCTAGAAGA	6900
AACTAATGAG GGATATGCTA TTGCGAAAAT TGCTGTAGCA AAAGCATGCG AATATATATC	6960
AAGAGAAAAC TCTAATTATT TTTATAAAAC AATTATCCA TGTAATTAT ATGGGAAATA	7020
TGATAAAATT GATGATAACT CGTCACATAT GATTCCGGCA GTTATAAAA AAATCCATCA	7080
TGCGAAAATT AATAATGTCC CAGAGATCGA AATTGGGGG GATGTAATT CGCGCCGTGA	7140
GTGTTATGTAT GCAGAAGATT TAGCTGATCT TATTTTTAT GTTATTCTTA AAATAGAATT	7200
CATGCCTAAT ATGGTAAATG CTGGTTAGG TTACGATTAT TCAATTAATG ACTATTATAA	7260
GATAATTGCA GAAGAAATTG GTTACTGAG GAGTTTTCT CATGATTAA CAAAACCAAC	7320
AGGAATGAAA CGGAAGCTAG TAGATATTTC ATTGCTTAAT AAAATTGGTT GGTCAAGTCA	7380
CTTTGAACTC AGAGATGGCA TCAGAAAGAC CTATAATTAT TACTGGAGA ATCAAAATAA	7440
ATGATTACAT ACCCACTTGC TAGTAATACT TGGGATGAAT ATGAGTATGC AGCAATACAG	7500
TCAGTAATTG ACTCAAAAAT GTTTACCATG GGTAAAAGG TTGAGTTATA TGAGAAAAT	7560
TTTGCTGATT TGTTGGTAG CAAATATGCC GTAATGGTTA GCTCTGGTTC TACAGCTAAT	7620
CTGTTAATGA TTGCTGCCCT TTTCTTCACT AATAAACCAA AACTTAAAAG AGGTGATGAA	7680
ATAATAGTAC CTGCAGTGTGTC ATGGTCTACG ACATATTACC CTCTGCAACA GTATGGCTTA	7740
AAGGTGAAGT TTGTCGATAT CAATAAGAA ACTTTAAATA TTGATATCGA TAGTTGAAA	7800
AATGCTATTT CAGATAAAAC AAAAGCAATA TTGACAGTAA ATTTATTAGG TAATCCTAAT	7860
GATTTTGCAA AAATAATGA GATAATAAT AATAGGGATA TTATCTTACT AGAAGATAAC	7920
TGTGAGTCGA TGGCGCGGT CTTCAAAAT AAGCAGGCAG GCACATTGG AGTTATGGGT	7980
ACCTTAGTT CTTTTACTC TCATCATATA GCTACAATGG AAGGGGGCTG CGTAGTTACT	8040
GATGATGAAG AGCTGTATCA TGTATTGGT TGCCTTCGAG CTCATGGTTG GACAAGAAAT	8100
TTACCAAAAG AGAATATGGT TACAGGCAGT AAGAGTGTATG ATATTTCGA AGAGTCGTT	8160
AAGTTTGTG TACAGGAGA CAATGTTCGC CCACCTGAAA TGAGTGGTGC TATTGGGATA	8220
GAGCAACTTA AAAAGTTACC AGGTTTATA TCCACCAGAC GTTCCAATGC ACAATATTT	8280
GTAGATAAAAT TTAAAGATCA TCCATTCTT GATATACAAA AAGAAGTTGG TGAAAGTAGC	8340

WO 98/50531

- 66 -

TGGTTGGTT	TTTCCTTCGT	TATAAAGGAG	GGAGCTGCTA	TTGAGAGGAA	GAGTTTAGTA	8400
AATAATCTGA	TCTCAGCAGG	CATTGAATGC	CGACCAATTG	TTACTGGAA	TTTCTCAA	8460
AATGAACGTG	TTTGAGTTA	TTTGATTAC	TCTGTACATG	ATACGGTAGC	AAATGCCGAA	8520
TATATAGATA	AGAATGGTT	TTTGTCGGA	AACCACCAGA	TACCTTGTT	TAATGAAATA	8580
GATTATCTAC	GAAAAGTATT	AAAATAACTA	ACGAGGCACT	CTATTCGAA	TAGAGTGCCT	8640
TTAAGATGGT	ATTAACAGTG	AAAAAAATTT	TAGCGTTGG	CTATTCTAAA	GTACTACCAC	8700
CGGTTATTGA	ACAGTTGTC	AATCCAATT	GCATCTTCAT	TATCACACCA	CTAATACTCA	8760
ACCACCTGGG	TAAGCAAAGC	TATGGTAATT	GGATTTATT	AATTACTATT	GTATCTTTT	8820
CTCAGTTAAT	ATGTGGAGGA	TGTTCCGCAT	GGATTGCAAA	AATCATTGCA	GAACAGAGAA	8880
TTCTTAGTGA	TTTATCAAAA	AAAAATGCTT	TACGTCAAAT	TTCCCTATAAT	TTTCAATTG	8940
TTATTATCGC	ATTGCGGTA	TTGATTCTT	TTCTTATATT	AAGTATTGT	TTCTTCGATG	9000
TTGCGAGGAA	TAATTCTCA	TTCTTATTG	CGATTATTAT	TTGTGGTTT	TTTCAGGAAG	9060
TTGATAATT	ATTTAGTGGT	GCGCTAAAAG	GTTTGAAAA	ATTTAATGTA	TCATGTTTT	9120
TTGAAAGTAAT	TACAAGAGTG	CTCTGGCCT	CTATAGTAAT	ATATGGCATT	TACGGAAATG	9180
CACTCTTATA	TTTTACATGT	TTAGCCTTA	CCATTAAAGG	TATGCTAAA	TATATTCTTG	9240
TATGCTGAA	TATTACCGGT	TGTTTCATCA	ATCCTAATT	TAATAGAGTT	GGGATTGTTA	9300
ATTTGTTAAA	TGAGTCAAA	TGGATGTTTC	TTCAATTAAAC	TGGTGGCGTC	TCACTTAGTT	9360
TTTTGATAG	GCTCGTAATA	CCATTGATT	TATCTGTAG	AAAACGGCT	TCTTATGTCC	9420
CTTGCCTTCA	ACTAGCTAA	TTGATGTTCA	CTCTTCTGC	GTCTGCAAAT	CAAATATTAC	9480
TACCAATGTT	TGCTAGAATG	AAAGCATCTA	ACACATTTC	CTCTAATTGT	TTTTTAAAAA	9540
TTCTGCTTGT	ATCACTAATT	TCTGTTTGC	CTTGTCTTG	GTTATTCTT	TTGGTCGTG	9600
ATATATTATC	AATATGGATA	AACCCTACAT	TTGCAACTGA	AAATTATAAA	TTAATGCAA	9660
TTTTAGCTAT	AAGTTACATT	TTATTGTCAA	TGATGACATC	TTTCATTTC	TTGTTATTAG	9720
GAATTGGTAA	ATCTAAGCTT	GTTGCAAATT	TAAATCTGGT	TGCAGGGCTC	GCACTTGCTG	9780
CTTCAACGTT	AATCGCAGCT	CATTATGCC	TTTATGCAAT	ATCTATGGTA	AAAATAATAT	9840
ATCCGGCTTT	TCAATTAT	TACCTTATG	TAGCTTTGT	CTATTTAAT	AGAGCGAAAA	9900
ATGTCTATTG	ATTTACTTTT	TTCAATTACT	GAAATCGCAA	TTGTTTTTC	TTGCACTATT	9960
TACATATTAA	CTCAATGTTT	GTAAATGCGG	AGGATCTATT	TAGATAAAAG	TATTTAATT	10020
CTTTTATGCT	TGCTCTTTT	TTTAGTAATC	ATTCAACTTC	CTGAGCTTAA	TGTAAACGGT	10080
TTGGTCGATT	CTTTAAAGTT	ATCACTGCCT	TTATTGATGG	TCTTATCGC	TTTCAAAAAA	10140
CCGAAATTAT	GCTTGTGGGT	TATTATTGCA	TTGTTGTTT	TGAACCTCTGC	ATTTAATT	10200
TTATATTAA	AGACATTGCA	TAAGTTTAGC	TCATTTCTT	TTACTTTTT	TATATTGCTG	10260
TTTTACTTGT	TTAGATTGGG	AATTGGTAAT	TTACCGGTTT	ATAAAAATAA	AAAATTTCAC	10320
GCGTGATTT	TTCTCTTAT	ATTAATAGAC	ATAATGCAGT	CATTGTTAAT	AAATTATAGG	10380

- 67 -

GGGCAGATTT TATATTCCGT AATTTGCATC	CTGATACTTG	TGTTTAAAGT	TAATTAAAGA	10440		
AAAAAGATTC CATACTTTTT	TTTAATGCTG	CCAGTTTAT	ATGTAATTAT	TATGGCTTAT	10500	
ATTGGTTTA ATTATTCAA	TAAAGCGTA	ACTTTTTTG	AACCTACAGC	AAGTAATATT	10560	
GAACGTACGG	GGATGATATA	TTATTTGGTT	TCACAGCTTG	GTGATTATAT	ATTCCATGGT	10620
ATGGGGACAT TAAATTCTT	AAATAACGGC	GGACAATATA	AGACGTTATA	TGGACTTCCA	10680	
TCATTAATTC	CTAATGACCC	TCATGATTTT	TTATTACGGT	TCTTATAAG	TATTGGTGTG	10740
ATAGGAGCAT	TGGTTTATCA	TTCTATATTT	TTTGTTTTTT	TTAGGAGAAT	ATCTTCTTA	10800
TTATATGAGA	GAAATGCTCC	TTTCATTGTT	GTAAGTTGTT	TGTTACTGTT	ACAAGTTGTG	10860
TTAATTATA	CATTAACCC	TTTGATGCT	TTTAATCGAT	TGATTGCGG	GCTTACAGTT	10920
GGAGTTGTT	ATGGATTGC	AAAAATTAGA	TAAGTATACC	TGTAATGGAA	ATTTAGACGC	10980
TCCACTTGT	TCAATAATCA	TTGCAACTTA	TAATTCTGAA	CTTGATATAG	CTAAGTGT	11040
GCAATCGGT	ACTAATCAAT	CTTATAAGAA	TATTGAAATC	ATAATAATGG	ATGGAGGATC	11100
TTCTGATAAA	ACGCTTGATA	TTGCAAAATC	GTAAAGAC	GACCGAATAA	AAATAGTTTC	11160
AGAGAAAGAT	CGTGGAAATT	ATGATGCCTG	GAATAAGCA	GTTGATTAT	CCATTGGTGA	11220
TTGGGTAGCA	TTTATTGGTT	CAGATGATGT	TTACTATCAT	ACAGATGCAA	TTGCTTCATT	11280
GATGAAGGGG	GTTATGGTAT	CTAATGGCGC	CCCTGTGGTT	TATGGGAGGA	CAGCGCACGA	11340
AGGTCCCGAT	AGGAACATAT	CTGGATTTC	AGGCAGTGAA	TGGTACAACC	TAACAGGATT	11400
TAAGTTTAAT	TATTACAAAT	GTAATTACC	ATTGCCATT	ATGAGCGCAA	TATATTCTCG	11460
TGATTCTTC	AGAACGAAAC	GTAAAGATAT	TAAATTAAAA	ATTGTTGCTG	ACGCTGATTG	11520
GTTTCTGAGA	TGTTTCATCA	AATGGAGTAA	AGAGAAGTCA	CCTTATTTA	TTAATGACAC	11580
GACCCCTATT	GTTAGAATGG	GATATGGTGG	GGTTTCGACT	GATATTCTT	CTCAAGTTAA	11640
AACTACGCTA	GAAAGTTCA	TTGTACGCAA	AAAGAATAAT	ATATCCTGTT	TAAACATACA	11700
GCTGATTCTT	AGATATGCTA	AAATTCTGGT	GATGGTAGCG	ATCAAAAATA	TTTTTGGCAA	11760
TAATGTTAT	AAATTAATGC	ATAACGGTA	TCATTCCCTA	AAGAAAATCA	AGAATAAAAT	11820
ATGAAGATTG	TTTATATAAT	AAACGGCTT	ACTTGTGGTG	GAGCCGAACA	CCTTATGACG	11880
CAGTTAGCAG	ACCAAATGTT	TATACCGGGG	CATGATGTTA	ATATTATTTG	TCTAACTGGT	11940
ATATCTGAGG	TAAAGCCAAC	ACAAAATATT	AATATTCTT	ATGTTAATAT	GGATAAAAAT	12000
TTTAGAAGCT	TTTTTAGAGC	TTTATTCAA	GTAAAAAAA	TAATTGTCGC	CTTAAAGCCA	12060
GATATAATAC	ATAGTCATAT	GTTTCATGCT	AATATTCTT	GTCGTTTAT	TAGGATGCTG	12120
ATTCCAGCGG	TGCCCCGTAT	ATGTACCGCA	CACAACAAA	ATGAAGGTGG	CAATGCAAGG	12180
ATGTTTGTT	ATCGACTGAG	TGATTCTTAA	GCTTCTTAA	CTACAAATGT	AAGTAAAGAG	12240
GCTGTTCAAG	AGTTTATAGC	AAGAAAGGCT	ACACCTAAA	ATAAAATAGT	AGAGATTCCG	12300
AATTTTATTA	ATACAAATAA	ATTTGATTTT	GATATTAATG	TCAGAAAGAA	AACGCGAGAT	12360
GCTTTAATT	TGAAAGACAG	TACAGCAGTA	CTGCTCGCAG	TAGGAAGACT	TGTTGAAGCA	12420

AAAGACTATC CGAACTTATT AAATGCAATA AATCATTGA TTCTTCAAA AACATCAAAT	12480
TGTAATGATT TTATTTGCT TATTGCTGGC GATGGCGCAT TAAGAAATAA ATTATTGGAT	12540
TTGGTTTGTCA ATTGAATCT TGTGGATAAA GTTTCTTCT TGGGGCAAAG AAGTGTATT	12600
AAAGAATTAA TGTGTGCTGC AGATCTTTT GTTTGAGTT CTGAGTGGGA AGGTTTTGGT	12660
CTCGTTGTTG CAGAAGCTAT GGCGTGTGAA CGTCCCGTTG TTGCTACCGA TTCTGGTGGA	12720
GTTAAAGAAG TC GTTGGACC TCATAATGAT GTTATCCCTG TCAGTAATCA TATTCTGTTG	12780
GCAGAGAAAA TCGCTGAGAC ACTTAAAATA GATGATAACG CAAGAAAAAT AATAGGTATG	12840
AAAAATAGAG AATATATTGT TTCCAATTTC TCAATTAAA CGATAGTGGAG TGAGTGGGAG	12900
CGCTTATATT TTAAATATTC CAAGCGTAAT AATATAATTG ATTGAAAATA TAAGTTGTA	12960
CTCTGGATGC AATAGTTCT CTATGCTGTT TTTTACTGG CTCCGTATTT TTACTTATAG	13020
CTGGATTTC TTATATATCA GTATTAATCT GTCTCAACTT CATCTAGACT ACATTCAAGC	13080
CGCGCATGCG TCGCGCGGTG ACTACACCTG ACAGGAGTAT GTAATGTCCA AGCAACAGAT	13140
CGGCGTCGTC GGTATGGCAG TGATGGGCG CAACCTGGCG CTCAACATCG AAAGCCGCGG	13200
TTATACCGTC TCCATCTTCA ACCGCTCCCG CGAGAAAAGT GAAGAAGTTG TTGCCGAGAA	13260
CCCGGATAAG AAACCTGGTTC CTTATTACAC GGTGAAAGAG TTCGTCGAGT CTCTTGAAAC	13320
CCCACGTCGT ATCCTGTTAA TGGTAAAAGC AGGGGCGGG ACTGATGCTG CTATCGATTC	13380
CCTGAAGCCG TATCTGGATA AAGGCGACAT CATTATTGAT GGTGGCAACA CCTTCTTCCA	13440
GGACACTATC CGTCGTAACC GTGAACGTGTC CGCGGAAGGC TTTAACTTCA TCGGTACCGG	13500
CGTGTCCGGC GGTGAAGAGG GCGCCCTGAA AGGCCCATCT ATCATGCCAG GTGCCAGAA	13560
AGAACCGTAT GAGCTGGTTG CGCCTATCCT GACCAAGATT GCTGCGGTTG CTGAAGATGG	13620
CGAACCATGT ATAACCTACA TCGGTGCTGA CGGTGCGGGT CACTACGTGA AGATGGTGCA	13680
CAACGGTATC GAATATGGCG ATATGCAGCT GATTGCTGAA GCCTATTCTC TGCTTAAAGG	13740
CGGCCTTAAT CTGTCTAACG AAGAGCTGGC AACCACTTT ACCGAGTGGA ATGAAGGCGA	13800
GCTAAGTAGC TACCTGATTG ACATCACCAA AGACATCTTC ACCAAAAAAG ATGAAGAGGG	13860
TAAATACCTG GTTGTGATGTGA TCCTGGACGA AGCTGCGAAC AAAGGCACCG GTAAATGGAC	13920
CAGCCAGAGC TCTCTGGATC TGGGTGAACC GCTGTCGCTG ATCACCGAAT CCGTATTGCG	13980
TCGCTACATC TCTTCTCTGA AAGACCAGCG CATTGCGGCA TCTAAAGTGC TGTCTGGTCC	14040
GCAGGCTAAA CTGGCTGGTG ATAAAGCAGA GTTCGTTGAG AAAGTCCGTC GCGCGCTGTA	14100
CCTGGGTAAA ATCGTCTCTT ATGCCAAGG CTTCTCTCAA CTGCGTGGCG CGTCTGACGA	14160
ATACAACTGG GATCTGAACG ACGGCGAAAT CGCGAAGATC TTCCGCGCGG GCTGCATCAT	14220
TCGTGCGCAG TTCCTGCAGA AAATTACTGA CGCGTATGCT GAAAACAAAG GCATTGCTAA	14280
CCTGTTGCTG GCTCCGTACT TCAAAAATAT CGCTGATGAA TATCAGCAAG CGCTGCGTGA	14340
TGTAGTGGCT TATGCTGTGC AGAACGGTAT TCCGGTACCG ACCTTCTCTG CAGCGGTAGC	14400
CTACTACGAC AGCTACCGTT CTGCGGTACT GCCGGCTAAT CTGATTCAAGG CACAGCGTGA	14460

TTACTTCGGT GCGCACACGT ATAAACGCAC TGATAAAGAA GGTGTGTTCC ACACCG

14516

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14024 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(v) ORIGINAL SOURCE

(A) ORGANISM: Escherichia coli

(vi) Note that the first 19bp is from the primer used for the long PCR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GTAACCAAGG GCGGTACGTG CATAAATTT AATGCTTATC AAAACTATTA GCATTAAAAA	60
TATATAAGAA ATTCTCAAAT GAACAAAGAA ACCGTTTCAA TAATTATGCC CGTTTACAAT	120
GGGGCCAAAA CTATAATCTC ATCAGTAGAA TCAATTATAC ATCAATCTTA TCAAGATT	180
GTGTTGTATA TCATTGACGA TTGTCACCC GATGATACAT TTTCATTAAT CAAACAGTCGA	240
TACAAAAACA ATCAGAAAAT AAGAATATTG CGTAACAAGA CAAATTTAGG TGTTGCAGAA	300
AGTCGAAATT ATGGAATAGA AATGGCCACG GGGAAATATA TTTCCTTTTG TGATGCGGAT	360
GATTTGTGGC ACGAGAAAAA ATTAGAGCGT CAAATCGAAG TGTAAATAA TGAATGTGTA	420
GATGTGGTAT GTTCTAATTA TTATGTTATA GATAACAATA GAAATATTGT TGGCGAAGTT	480
AATGCTCCTC ATGTGATAAA TTATAGAAAA ATGCTCATGA AAAACTACAT AGGGAATTG	540
ACAGGAATCT ATAATGCCAA CAAATTGGGT AAGTTTATC AAAAAAGAT TGGTCACGAG	600
GATTATTTGA TGTGGCTGGA AATAATTAAT AAAACAAATG GTGCTATTG TATTCAAGAT	660
AATCTGGCGT ATTACATGCG TTCAAATAAT TCACTATCGG GTAATAAAAT TAAAGCTGCA	720
AAATGGACAT GGAGTATATA TAGAGAACAT TTACATTGTT CCTTTCCAAA AACATTATAT	780
TATTTTTAT TATATGCTTC AAATGGAGTC ATGAAAAAAA TAACACATTC ACTATTAAGG	840
AGAAAGGAGA CTAAGAAAGTG AAGTCAGCGG CTAAGTTGAT TTTTTTATTC CTATTTACAC	900
TTTATAGTCT CCAGTTGTAT GGGGTTATCA TAGATGATCG TATAACAAAT TTTGATAACAA	960
AGGTATTAAC TAGTATTATA ATTATATTTC AGATTTTTT TGTTTTATTA TTTTATCTAA	1020
CGATTATAAA TGAAAGAAAA CAGCAGAAAA AATTTATCGT GAACTGGGAG CTAAAGTTAA	1080
TACTCGTTT CCTTTTTGTG ACTATAGAAA TTGCTGCTGT AGTTTTATTT CTTAAAGAAG	1140

WO 98/50531

- 70 -

GTATTCCTAT	ATTTGATGAT	GATCCAGGGG	GGGCTAAACT	TAGAATAGCT	GAAGGTAATG	1200
GACTTTACAT	TAGATATATT	AAGTATTG	GTAATATAGT	TGTGTTGCA	TTAATTATTC	1260
TTTATGATGA	GCATAAATTC	AAACAGAGGA	CCATCATATT	TGTATATTT	ACAAACGATTG	1320
CTTTATTG	TTATCGTTCT	GAATTGGTGT	TGCTCATTCT	TCAATATATA	TTGATTACCA	1380
ATATCCTGTC	AAAGGATAAC	CGTAATCCTA	AAATAAAAAG	AATAATAGGG	TATTTTTTAT	1440
TGGTAGGGGT	TGTATGCTCG	TTGTTTATC	TAAGTTAGG	ACAAGACGGA	GAACAAAATG	1500
ACTCATATAA	TAATATGTTA	AGGATAATTA	ATAGGTTAAC	AATAGAGCAA	GTTGAAGGTG	1560
TTCCATATGT	TGTTTCTGAA	TCTATTAAGA	ACGATTTCTT	TCCGACACCA	GAGTTAGAAA	1620
AGGAATTAAA	AGCAATAATA	AATAGAATAC	AGGGAATAAA	GCATCAAGAC	TTATTTTATG	1680
GAGAACGGTT	ACATAAACAA	GTATTTGGAG	ACATGGGAGC	AAATTTTTA	TCAGTTACTA	1740
CGTATGGAGC	AGAACTGTTA	GTTTTTTTG	GTTCCTCTG	TGTATTCATT	ATCCCTTTAG	1800
GGATATATAT	ACCTTTTAT	CTTTAAAGA	GAATGAAAAA	AACCCATAGC	TCGATAAAATT	1860
GCGCATTCTA	TTCATATATC	ATTATGATT	TATTGCAATA	CTTAGTGGCT	GGGAATGCAT	1920
CGGCCTTCTT	TTTTGGTCCT	TTTCTCTCCG	TATTGATAAT	GTGTACTCCT	CTGATCTTAT	1980
TGCATGATAC	GTAAAGAGA	TTATCACGAA	ATGAAAATAT	CAGTTATAAC	TGTGACTTAT	2040
AATAATGCTG	AAGGGTTAGA	AAAAACTTTA	AGTAGTTTAT	CAATTTAAA	AATAAAACCT	2100
TTTGAGAGTA	TTATAGTTGA	TGGCGCTCT	ACAGATGGAA	CGAACCGTGT	CATTAGTAGA	2160
TTTACTAGTA	TGAATATTAC	ACATGTTTAT	AAAAAAGATG	AAGGGATATA	TGATGCGATG	2220
AATAAGGGCC	GAATGTTGGC	CAAAGGCAC	TTAACACATT	ATTTAACGC	CGGCGATAGC	2280
GTAATTGGAG	ATATATATAA	AAATATCAA	GAGCCATGTT	TGATTAAGT	TGGCCTTTTC	2340
GAAAATGATA	AACTTCTGGG	ATTTCTTCT	ATAACCCATT	CAAATACAGG	GTATTGTCAT	2400
CAAGGGGTGA	TTTCCCAAA	GAATCATTCA	GAATATGATC	TAAGGTATAA	AATATGTGCT	2460
GATTATAAGC	TTATTCAAGA	GGTGTTCCT	GAAGGGTTAA	GATCTCTATC	TTTGATTACT	2520
TCGGGTTATG	TAAAATATGA	TATGGGGGGA	GTATCTCAA	AAAAAAGAAT	TTTAAGAGAT	2580
AAAGAGCTTG	CCAAAATTAT	GTTCGAAAAA	AATAAAAAAA	ACCTTATTAA	GTTCATTCCA	2640
ATTTCAATAA	TCAAAATT	ATTCCCTGAA	CGTTTAAGAA	GAGTATTGCG	AAAATGCAA	2700
TATATTTGTC	TAACTTATT	CTTCATGAAG	AATAGTCAC	CATATGATAA	TGAATAAAAT	2760
CAAAAAAATA	CTTAAATT	GCAC	TTAA	AAAATATGAT	ACATCAAGTG	2820
AGAACAGGAA	AGGTACAGGA	TTATATCCTT	GTCTGTTATT	TCAAGTTGA	TTAGTAAAT	2880
ACTCTCACTA	CTTTCTCTTA	TATTA	ACTGT	AAGTTAACT	TTACCTTATT	2940
GAGATTTGGT	GTATGGATGA	CTATTACCA	TCTTGGTGCT	GCTCTGACAT	TTTGACTT	3000
AGGTATAGGA	AATGCATTAA	CAAACAGGAT	CGCACATTCA	TTTGCCTGTG	GCAAAATTT	3060
AAAGATGAGT	CGGCAAATTA	GTGGTGGGCT	CACTTGCTG	GCTGGATTAT	CGTTGTCAT	3120
AACTGCAATA	TGCTATATTA	CTTCTGGCAT	GATTGATTGG	CAACTAGTAA	AAAAAGGTAT	3180

- 71 -

AAACGAGAAT	GTGTATGCAG	AGTTACAACA	CTCAATTAAA	GTCTTGTAA	TCATATTTGG	3240
ACTTGGAAATT	TATTCAAATG	GTGTGCAAAA	AGTTTATATG	GGAATACAAA	AAGCCTATAT	3300
AAGTAATATT	GTAAATGCCA	TATTTATATT	GTTATCTATT	ATTACTCTAG	TAATATCGTC	3360
GAAACTACAT	GCGGGACTAC	CAGTTTAAT	TGTCAGCACT	CTTGGTATTC	AATACATATC	3420
GGGAATCTAT	TTAACAAATTA	ATCTTATTAT	AAAGCGATTA	ATAAAGTTA	CAAAGTTAA	3480
CATACATGCT	AAAAGAGAAG	CTCCATATTT	GATATTAAAC	GGTTTTTCT	TTTTTATTTT	3540
ACAGTTAGGC	ACTCTGGCAA	CATGGAGTGG	TGATAACTTT	ATAATATCTA	TAACATTGGG	3600
TGTTACTTAT	GTTGCTGTTT	TTAGCATTAC	ACAGAGATTA	TTTCAAATAT	CTACGGTCCC	3660
TCTTACGATT	TATAACATCC	CGTTATGGGC	TGTTATGCA	GATGCTCATG	CACGCAATGA	3720
TACTCAATTT	ATAAAAAAAGA	CGCTCAGAAC	ATCATTGAAA	ATAGTGGTA	TTTCATCATT	3780
CTTATTGGCC	TTCATATTAG	TAGTGTTCGG	TAGTGAAGTC	GTTAATATTT	GGACAGAAGG	3840
AAAGATTCA	GTACCTCGAA	CATTCTATAAT	AGCTTATGCT	TTATGGTCTG	TTATTGATGC	3900
TTTTTCAAT	ACATTTGCAA	GCTTTTAAA	TGGTTGAAC	ATAGTTAAC	AACAAATGCT	3960
TGCTGTTGTA	ACATTGATAT	TGATCGCAAT	TCCAGCAAAA	TACATCATAG	TTAGCCATT	4020
TGGGTTAACT	GTTATGTTGT	ACTGCTTCAT	TTTTATATAT	ATTGTAAATT	ACTTTATATG	4080
GTATAAATGT	AGTTTAAAGA	AACATATCGA	TAGACAGTTA	AATATAAGAG	GATGAAAATG	4140
AAATATATAC	CAGTTTACCA	ACCGTCATTG	ACAGGAAAAG	AAAAAGAATA	TGTAAATGAA	4200
TGTCTGGACT	CAACGTGGAT	TTCATCAAAA	GGAAACTATA	TTCAGAAGTT	TGAAAATAAA	4260
TTTGCAGAAC	AAAACCATGT	GCAATATGCA	ACTACTGTAA	GTAATGGAAC	GGTTGCTCTT	4320
CATTTAGCTT	TGTTAGCGTT	AGGTATATCG	GAAGGAGATG	AAGTTATTGT	TCCAACACTG	4380
ACATATATAG	CATCAGTTAA	TGCTATAAAA	TACACAGGAG	CCACCCCCAT	TTTCGTTGAT	4440
TCAGATAATG	AAACTTGGCA	AATGCTGTT	AGTGACATAG	AACAAAAAAT	CACTAATAAA	4500
ACTAAAGCTA	TTATGTTGTT	CCATTATAC	GGACATCCAT	GTGATATGGA	ACAAATTGTA	4560
GAACCTGGCCA	AAAGTAGAAA	TTTGTGTTGTA	ATTGAAGATT	GCGCTGAAGC	CTTTGGTTCT	4620
AAATATAAAG	GTAATATGT	GGGAACATT	GGAGATATT	CTACTTTAG	CTTTTTGGA	4680
AATAAAACTA	TTACTACAGG	TGAAGGTGGA	ATGGTTGTCA	CGAATGACAA	AACACTTTAT	4740
GACCGTTGTT	TACATTTAA	AGGCCAAGGA	TTAGCTGTAC	ATAGGCAATA	TTGGCATGAC	4800
GTTATAGGCT	ACAATTATAG	GATGACAAAT	ATCTGCGCTG	CTATAGGATT	AGCCCAGTTA	4860
GAACAAAGCTG	ATGATTTAT	ATCACGAAAA	CGTGAAATTG	CTGATATTTA	AAAAAAAAT	4920
ATCAACAGTC	TTGTACAAGT	CCACAAGGAA	AGTAAAGATG	TTTTTCACAC	TTATTGGATG	4980
GTCTCAATTG	TAACTAGGAC	CGCAGAGGAA	AGAGAGGAAT	TAAGGAATCA	CCTTGCAGAT	5040
AAACTCATCG	AAACAAGGCC	AGTTTTTAC	CCTGTCCACA	CGATGCCAAT	GTACTCGGAA	5100
AAATATCAAA	AGCACCCAT	AGCTGAGGAT	CTTGGTTGGC	GTGGAATTAA	TTTACCTAGT	5160
TTCCCCAGCC	TATCGAATGA	GCAAGTTATT	TATATTGTTG	AATCTATTAA	CGAATTAT	5220

- 72 -

AGTGATAAAAT AGCCTAAAAT ATTGTAAAGG TCATTCATGA AAATTGCGTT GAATTCAGAT	5280
GGATTTACG AGTGGGGCGG TGGATTGAT TTTATTAAAT ATATTCTGTC AATATTAGAA	5340
ACGAAACCAG AAATATGTAT CGATATTCTT TTACCGAGAA ATGATATACA TTCTCTTATA	5400
AGAGAAAAAG CATTCCCTT TAAAAGTATA TTAAAAGCAA TTTTAAAGAG GGAAAGGCCT	5460
CGATGGATT CATTAAATAG ATTTAATGAG CAATACTATA GAGATGCCCT TACACAAAAT	5520
AATATAGAGA CGAATCTTAC CTTTATTAAA AGTAAGAGCT CTGCCTTTA TTCATATTTT	5580
GATAGTAGCG ATTGTGATGT TATTCTCCT TGCAATGCGTG TTCCCTCGGG AAATTTGAAT	5640
AAAAAAGCAT GGATTGGTTA TATTTATGAC TTTCAACACT GTTACTATCC TTCATTTTT	5700
AGTAAGCGAG AAATAGATCA AAGGAATGTG TTTTTAAAT TGATGCTCAA TTGCGCTAAC	5760
AATATTATTG TTAATGCACA TTCAGTTATT ACCGATGCAA ATAAATATGT TGGGAATTAT	5820
TCTGCAAAAC TACATTCTCT TCCATTAGT CCATGCCCTC AATTAAAATG GTTCGCTGAT	5880
TACTCTGGTA ATATTGCCAA ATATAATATT GACAAGGATT ATTTTATAAT TTGCAATCAA	5940
TTTTGGAAAC ATAAAGATCA TGCAACTGCT TTTAGGGCAT TTAAAATTAA TACTGAATAT	6000
AATCCTGATG TTTATTTAGT ATGCACGGGA GCTACTCAAG ATTATCGATT CCCTGGATAT	6060
TTTAATGAAT TGATGGTTT GGCAAAAAAG CTCGGAATTG AATCGAAAAT TAAGATATTA	6120
GGGCATATAC CTAAACATTGA ACAAAATTGAA TTAATCAAAA ATTGCATTGC TGTAATACAA	6180
CCAACCTTAT TTGAAGGCAG GCCTGGAGGG GGGGTAACAT TTGACGCTAT TGCATTAGGG	6240
AAAAAAGTTA TACTATCTGA CATAGATGTC AATAAAGAAG TTAATTGCGG TGATGTATAT	6300
TTCTTTCAGG CAAAAAACCA TTATTCTTA AATGACGCGA TGGTAAAAGC TGATGAATCT	6360
AAAATTTTT ATGAACCTAC AACTCTGATA GAATTGGTC TCAAAAGACG CAATGCGTGT	6420
GCAGATTTC TTTTAGATGT TGTGAAACAA GAAATTGAAT CCCGATCTTA ATATATTCAA	6480
GAGGTATATA ATGACTAAAG TCGCTCTTAT TACAGGTGTA ACTGGACAAG ATGGATCTTA	6540
TCTAGCTGAG TTTTGCTTG ATAAAGGGTA TGAAGTTCAT GGTATCAAAC GCCGAGCCTC	6600
ATCTTTAAT ACAGAACGCA TAGACCATAT TTATCAAGAT CCACATGGTT CTAACCCAAA	6660
TTTTCACTTG CACTATGGAG ATCTGACTGA TTCATCTAAC CTCACTAGAA TTCTAAAGGA	6720
GGTACAGCCA GATGAAGTAT ATAATTTAGC TGCTATGAGT CACGTAGCAG TTTCTTTGA	6780
GTCTCCAGAA TATACAGCCG ATGTCGATGC AATTGGTACA TTACGTTAC TGGAAGCAAT	6840
TCGCTTTTA GGATTGGAAA ACAAAACGCG TTTCTATCAA GCTTCAACCT CAGAATTATA	6900
TGGACTTGTT CAGGAAATCC CTCAAAAAGA ATCCACCCCT TTTTATCCTC GTTCCCCTTA	6960
TGCAGTTGCA AAACTTTACG CATATTGGAT CACGGTAAAT TATCGAGAGT CATATGGTAT	7020
TTATGCATGT AATGGTATAT TGTTCAATCA TGAATCTCCA CGCCGTGGAG AAACGTTTGT	7080
AACAAGGAAA ATTACTCGAG GACTTGCAA TATTGCACAA GGCTTGGAAAT CATGTTTGT	7140
TTTAGGAAAT ATGGATTCGT TACGAGATTG GGGACATGCA AAAGATTATG TTAGAATGCA	7200
ATGGTTGATG TTACAACAGG AGCAACCCGA AGATTTGTG ATTGCAACAG GAGTCCAATA	7260

- 73 -

CTCAGTCGGT CAGTTGTCG AAATGGCAGC AGCACAACTT GGTATTAAGA TGAGCTTTGT	7320
TGGTAAAGGA ATCGAAGAAA AAGGCATTGT AGATTCGGTT GAAGGACAGG ATGCTCCAGG	7380
TGTGAAACCA GGTGATGTCA TTGTTGCTGT TGATCCTCGT TATTTCCGAC CAGCTGAAGT	7440
TGATACTTTG CTTGGAGATC CGAGCAAAGC TAATCTAAA CTTGGTTGGA GACCAGAAAT	7500
TACTCTTGC GAAATGATT CTGAAATGGT TGCCAAAGAT CTTGAAGCCG CTAAAAAACAA	7560
TTCTCTTTA AAATCGCATG GTTTTCTGT AAGCTTAGCT CTGGAATGAT GATGAATAAG	7620
CAACGTATTT TTATTGCTGG TCACCAAGGA ATGGTTGGAT CAGCTATTAC CCGACGCC	7680
AAACAACGTG ATGATGTTGA GTTGGTTTA CGTACTCGGG ATGAATTGAA CTTGGTGGAT	7740
AGTAGCGCTG TTTGGATT TTTTCTTCAG CAGAAAATCG ACCAGGTTA TTTGGCAGCA	7800
GCAAAAGTCG GAGGTATTT AGCTAACAGT TCTTATCCTG CCGATTTAT ATATGAGAAT	7860
ATAATGATAG AGGCGAATGT CATTGATGCT GCCCACAAAA ATAATGTAAA TAAACTGCTT	7920
TTCCTCGGTT CGTCGTGTAT TTATCCTAAG TTAGCACACC AACCGATTAT GGAAGACGAA	7980
TTATTACAAG GGAAACATTGA GCCAACAAAT GAACCTTATG CTATCGAAA AATTGCGAGT	8040
ATTAATTAT GTGAATCTTA TAACCGTCAG TTTGGCGTG ATTACCGTTC AGTAATGCCA	8100
ACCAATCTT ATGGTCCAAA TGACAATTCTT CATCCAAGTA ATTCTCATGT GATTCCGGCG	8160
CTTTTGCGCC GCTTTCATGA TGCTGTGGAA ACAATTCTC CGAATGTTGT TGTTTGGGGA	8220
AGTGGTACTC CAAAGCGTGA ATTCTTACAT GTAGATGATA TGGCTCTGC AAGCATTAT	8280
GTCATGGAGA TGCCATACGA TATATGGCAA AAAAATACTA AAGTAATGTT GTCTCATATC	8340
AATATTGGAA CAGGTATTGA CTGCACGATT TGTGAGCTTG CGGAAACAAT AGCAAAAGTT	8400
GTAGGTTATA AAGGGCATAT TACGTTCGAT ACAACAAAGC CCGATGGAGC CCCTCGAAA	8460
CTACTTGATG TAACGTTCT TCATCAACTA GGTTGGAATC ATAAAATTAC CCTTCACAAG	8520
GGTCTTGAAA ATACATACAA CTGGTTCTT GAAAACCAAC TTCAATATCG GGGGTAATAA	8580
TGTTTTACA TTCCCAAGAC TTTGCCACAA TTGTAAGGTC TACTCCTCTT ATTTCTATAG	8640
ATTTGATTGT GGAAAACGAG TTTGGCGAAA TTTGCTAGG AAAACGAATC AACCGCCCG	8700
CACAGGGCTA TTGGTTCGTT CCTGGTGGTA GGGTGTGAA AGATGAAAAA TTGCAGACAG	8760
CCTTTGAACG ATTGACAGAA ATTGAACTAG GAATTCGTTT GCCTCTCTCT GTGGGTAAGT	8820
TTTATGGTAT CTGGCAGCAC TTCTACGAAG ACAATAGTAT GGGGGGAGAC TTTCAACGC	8880
ATTATATAGT TATAGCATTC CTTCTTAAAT TACAACCAA CATTGAAA TTACCGAAGT	8940
CACAACATAA TGCTTATTGC TGGCTATCGC GAGCAAAGCT GATAAATGAT GACGATGTGC	9000
ATTATAATTG TCGCGCATAT TTTAACAAATA AAACAAATGA TCGCATTGGC TTAGATAATA	9060
AGGATATAAT ATGTCTGATG CGCCAATAAT TGCTGTAGTT ATGGCCGGTG GTACAGGCAG	9120
TCGTCTTGG CCACCTTCTC GTGAACATAA TCCAAAGCAG TTTTACAAC TCTCTGGTGA	9180
TAACACCTTG TTACAAACGA CTTGCTACG ACTTTCAAGGC CTATCATGTC AAAACCATT	9240
AGTGATAACA AATGAACAGC ATCGCTTTGT TGTGGCTGAA CAGTTAAGGG AAATAAATAA	9300

- 74 -

ATTAATGGT AATATTATTC TAGAACCATG. CGGGCGAAAT ACTGCACCAAG CAATAGCGAT	9360
ATCTCGTTT CATCGTTAA AACGTAATCC TCAGGAAGAT CCATTGCTTC TAGTTCTTGC	9420
GGCAGACCAAC GTTATAGCTA AAGAAAGTGT TTTCTGTGAT GCTATTAAAA ATGCAACTCC	9480
CATCGCTAAT CAAGGTAAAA TTGTAACGTT TGGAATTATA CCAGAATATG CTGAAACTGG	9540
TTATGGGTAT ATTGAGAGAG GTGAACTATC TGTACCGCTT CAAGGGCATG AAAATACTGG	9600
TTTTTATTAT GTAAATAAGT TTGTCGAAAA GCCTAATCGT GAAACCGCAG AATTGTATAT	9660
GACTTCTGGT AATCACTATT GGAATAGTGG AATATTCATG TTTAAGGCAT CTGTTTATCT	9720
TGAGGAATTG AGAAAATTAA GACCTGACAT TTACAATGTT TGTGAACAGG TTGCCTCATC	9780
CTCATACATT GATCTAGATT TTATTGATT ATCAAAAGAA CAATTCAAG ATTGTCCCTGC	9840
TGAATCTATT GATTTGCTG TAATGGAAAA AACAGAAAAA TGTGTTGTAT GCCCTGTTGA	9900
TATTGGTTGG AGTGACGTTG GATCTGGCA ATCGTTATGG GACATTAGTC TAAAATCGAA	9960
AACAGGAGAT GTATGTAAAG GTGATATATT AACCTATGAT ACTAAGAATA ATTATATCTA	10020
CTCTGAGTCA CGCTTGGTAG CCGCCATTGG AATTGAAGAT ATGGTTATCG TGCAAACATA	10080
AGATGCCGTT CTTGTGTCTA AAAAGAGTGA TGTACAGCAT GTAAAAAAA TAGTCGAAAT	10140
GCTTAAATTG CAGCAACGTA CAGAGTATAT TAGTCATCGT GAAGTTTCC GACCATGGG	10200
AAAATTGAT TCGATTGACC AAGGTGAGCG ATACAAAGTC AAGAAAATTA TTGTGAAACC	10260
TGGTGAGGGG CTTTCTTTAA GGATGCATCA CCATCGTTCT GAACATTGGA TCGTGCCTTC	10320
TGGTACAGCA AAAGTAACCC TTGGCGATAA AACTAAACTA GTCACCGCAA ATGAATCGAT	10380
ATACATTCCC CTTGGCGCAG CGTATAGTCT TGAGAATCCG GGCATAATCC CTCTTAATCT	10440
TATTGAAGTC AGTTCAGGGG ATTATTTGGG AGAGGATGAT ATTATAAGAC AGAAAGAACG	10500
TTACAAACAT GAAGATTAAC ATATGAAATC TTTAACCTGC TTTAAAGCCT ATGATATTCG	10560
CGGGAAATTAA GGCGAAGAAC TGAATGAAGA TATTGCCTGG CGCATTGGC GTGCCTATGG	10620
CGAATTTCCTC AAACCGAAAA CCATTGTTT AGGCAGGTGAT GTCCGCCTCA CCAGCGAACG	10680
GTAAAACGT GCGCTTGCAG AAGGTTACA GGATGCGGGC GTCGATGTGC TGGATATCGG	10740
TATGTCGGC ACCGAAGAGA TCTATTCGC CACGTTCCAT CTGGAGTGG ATGGCGGCAT	10800
CGAAGTTACC GCCAGCCATA ACCCGATGGA TTACAACGGC ATGAAGCTGG TGCGCGAAGG	10860
GGCTCGCCCG ATCAGCGGTG ATACCGGACT GCGCGATGTC CAGCGTCTGG CAGAAGCCAA	10920
TGACTTCCCT CCTGTCGATG AAACCAAACG TGGTCGCTAT CAGCAAATCA ATCTGCGTGA	10980
CGCTTACGTT GATCACCTGT TCGGTTATAT CAACGTCAA AACCTCACGC CGCTCAAGCT	11040
GGTGATCAAC TCCGGGAACG GCGCAGCGGG TCCGGTGGTG GACGCCATTG AAGCCCGATT	11100
TAAAGCCCTC GGCGCACCGG TGGATTAAAT CAAAGTACAC AACACGCCGG ACGGCAATTT	11160
CCCCAACGGT ATTCTAACCG CGCTGCTGCC GGAATGCCGC GACGACACCC GTAATGCGGT	11220
CATCAAACAC GGCGCGGATA TGGCATTGC CTTTGATGGC GATTTTGACC GCTGTTCCCT	11280
GTGGACGAA AAAGGGCAAGT TTATCGAGGG CTACTACATT GTCGGCCTGC TGGCAGAACG	11340

GTTCCCTCGAA	AAAAATCCCG	GCGCGAAGAT	CATCCACGAT	CCACGTCTCT	CCTGGAACAC	11400
CGTTGATGTG	GTGACTGCCG	CAGGCGGCAC	CCCGGTAATG	TCGAAAACCG	GACACGCCTT	11460
TATTAAAGAA	CGTATGCGCA	AGGAAGACGC	CATCTACGGT	GGCGAAATGA	GCGCTCACCA	11520
TTACTTCCGT	GATTCGCTT	ACTGCGACAG	CGGCATGATC	CCGTGGCTGC	TGGTCGCCGA	11580
ACTGGTGTGC	CTGAAAGGAA	AAACGCTGGG	CGAAATGGTG	CGCGACCGGA	TGGCGCGTT	11640
TCCGGCAAGC	GGTGAGATCA	ACAGCAAACCT	GGCGCAACCC	GTTGAGGCAA	TTAATCGCGT	11700
GGAACAGCAT	TTTAGCCGCG	AGGCGCTGGC	GGTGGATCGC	ACCGATGGCA	TCAGCATGAC	11760
CTTGGCCGAC	TGGCGCTTTA	ACCTGCGCTC	CTCCAACACC	GAACCGGTGG	TGGCGTTGAA	11820
TGTGGAATCA	CGCGGTGATG	TAAAGCTAAT	GGAAAAGAAA	ACTAAAGCTC	TTCTTAAATT	11880
GCTAA GTGAG	TGATTATTTA	CATTAATCAT	TAAGCGTATT	TAAGATTATA	TTAAAGTAAT	11940
GTTATTGCGG	TATATGATGA	ATATGTGGC	TTTTTTATGT	ATAACGACTA	TACCGCAACT	12000
TTATCTAGGA	AAAGATTAAT	AGAAATAAAG	TTTTGTACTG	ACCAATTGTC	ATTTCACGTC	12060
ACGATTGAGA	CGTTCCCTTG	CTTAAGACAT	TTTTTCATCG	CTTATGTAAT	AACAAATGTG	12120
CCTTATATAA	AAAGGAGAAC	AAAATGGAAC	TTAAAATAAT	TGAGACAATA	GATTTTTATT	12180
ATCCCTGTTT	ACGGATATTAT	AGCCAAAGTT	GTATCCTGCA	TCAGTCCTGC	AATATTCAC	12240
GAGTGCTTTG	TTAACTGAAT	ACATGCTGTC	CATTTTCCAG	ATGATAACGA	CGTCATCGCA	12300
ATTGATGGTA	AAACACTTCG	GCACACTTAT	GACAAGAGTC	GTCGCAGAGG	AGTGGTTCAT	12360
GTCATTAGTG	CGTTTCAGCA	ATGCACAGTC	TGGTCCTCGG	ATAGATCAAG	ACGGATGAGA	12420
AACCTAATGC	GTTCACAGTT	ATTCAATGAA	TTTCTAAAAT	GATGGGTATT	AAAGGAAAAA	12480
TAATCATAAC	TGATGCGATG	GCTTGCCAGA	AAGATATTGC	AGAGAAGATA	AAAAAACAGA	12540
GATGTGATTA	TTTATTCGCT	GTAAAAGGAA	ATAAGAGTCG	GCTTAATAGA	GTCTTTGAGG	12600
AGATATTTAC	GCTGAAAGAA	TTAAATAATC	CAAAACATGA	CAGTTACGCA	ATTAGTGAAA	12660
AGAGGCACGG	CAGAGACGAT	GTCCGTCTTC	ATATTGTTG	AGATGCTCCT	GATGAGCTTA	12720
TTGATTCAC	TTTGAATGG	AAAGGGCTGC	AGAATTTATG	AATGGCAGTC	CACTTTCTCT	12780
CAATAATAGC	AGAGCAAAAG	AAAGAATCCG	AAATGACGAT	CAAATATTAT	ATTAGATCTG	12840
CTGCTTTAAC	CGCAGAGAAC	TTGCCACAG	TAAATCGAAA	TCACTGGCGC	ATGGAGAATA	12900
AGTTGCACAG	TAGCCTGATG	TGGTAATGAA	TGAAATCGAC	TATAATATAA	GAAGGCGAGT	12960
TGCATTGAA	TGATTTCTA	GAATGCGGCA	CATCGCTATT	AATATCTGAC	AATGATAATG	13020
TATTCAAGGC	AGGATTATCA	TGTAAGATGC	GAAAAGCACT	CATGGACAGA	AACTTCCTAG	13080
CGTCAGGCAT	TGCAGCGTGC	GGGCTTCAT	AATCTTGCAT	TGGTTTGAT	AAGATATTTC	13140
TTTGGAGATG	GGAAAATGAA	TTTGTATGGT	ATTTTGGTG	CTGGAAGTTA	TGGTAGAGAA	13200
ACAATACCCA	TTCTAAATCA	ACAAATAAAG	CAAGAATGTG	GTTCTGACTA	TGCTCTGGTT	13260
TTTGTGGATG	ATGTTTTGGC	AGGAAAGAAA	GTAAATGGTT	TTGAAGTGCT	TTCAACCAAC	13320
TGCTTTCTAA	AAGCCCTTA	TTTAAAAAAG	TATTTTAATG	TTGCTATTGC	TAATGATAAG	13380

- 76 -

ATACGACAGA GAGTGTCTGA GTCAATATTA TTACACGGGG TTGAACCAAT AACTATAAAA	13440
CATCCAAATA GCGTGTGTTA TGATCATACT ATGATAGGTA GTGGCGCTAT TATTTCTCCC	13500
TTTGTACAA TATCTACTAA TACTCATATA GGGAGGTTTT TTCATGAAA CATATACTCA	13560
TACGTTGCAC ATGATTGTCA AATAGGAGAC TATGTTACAT TTGCTCCTGG GGCTAAATGT	13620
AATGGATATG TTGTTATTGA AGACAATGCA TATATAGGCT CGGGTGCAGT AATTAAGCAG	13680
GGTGTTCCTA ATCGCCCCTGC TATTATTGGC GCGGGAGCCA TTATAGGTAT GGGGGCTGTT	13740
GTCACTAAA GTGTTCCGC CGGTATAACT GTGTGCGGAA ATCCAGCAAG AGAAATGAAA	13800
AGATCGCCAA CATCTATTAA ATGGGAATGC GAAAACACGT TCCAAATGGG ACTAATGTTT	13860
AAAATATATA TAATTCGCT AATTTACTAA ATTATGGCTT CTTTTTAAGC TATCCTTAC	13920
TTAGTTATTA CTGATACAGC ATGAAATTAA TAATACTCTG ATACATTTT ATACGTTATT	13980
CAAGCCGCAT ATCTAGCGGT AACCCCTGAC AGGAGTAAAC AATG	14024

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12441 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Salmonella enterica* serovar *muenchen* serogroup C2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GTTGACAAAT ACCGACCGTA TAATGAATCA AACGTTCTGG ATTGGTATTT ATCCAGGCTT	60
GACTACAGAG CATTAGATT ATGTCGTAAG TAAGTTGAA GAATTTTTG GTTTAAATTT	120
CTAATTTTA GGATAGGATG CTTGATGTGA ATAAGAAAAT CCTAATGACT GGCGCTACTA	180
GCTTTGAGG TACCCATCTA CTACATAGTC TCATAAAGGA AGGTTATAGT ATTATTGCAT	240
TAAAGCGTCC TATAACCGAG CCAACGATTA TCAATACCTT GATTGAATGG TTGAATATAC	300
AAGATATAGA AAAAATATGT CAATCATCTA TGAATATTCA TGCGATTGTC CATATTGCAA	360
CAGACTATGG TCGAACAGA ACCCTATAT CTGAACAATA TAAATGTAAT GTCCTATTAC	420
CAACAAGACT GCTTGAGTTA ATGCCAGCGC TTAAAACGAA ATTCTTTATT TCTACTGACT	480
CTTTTTTGG GAAATATGAG AAGCACTATG GATATATGCG TTCTTACATG GCATCTAAA	540
GACATTTGT AGAACTATCA AAAATATACG TAGAGGAACA TCCAGACGTT TGTTTTATAA	600
ATTTACGTTT AGAACATGTT TACGGTGAGA GGGATAAAGC AGGTAAAATA ATCCCGTATG	660
TTATCAAAAA AATGAAAAAC AATGAAGATA TTGATTGTAC GATCGCCAGG CAGAAAAGAG	720
ATTTTATTAA TATAGACGAT GTTGTTCGG CCTATTGAA AATTTAAAG GAGGGTTTA	780

- 77 -

ACGCTGGACA	CTATGATGTC	GAGGTGGGGA	CTGGAAAATC	GATAGAGCTA	AAAGAAGTGT	840
TTGAGATAAT	AAAAAAAGAA	ACGCATAGTA	GTAGTAAGAT	AAATTATGGT	GCAGTTGCGA	900
TGCGTGATGA	TGAGATTATG	GAGTCACATG	CAAATACCTC	TTTCTTGACT	CGATTAGGTT	960
GGAGTGCCGA	GTTCCTATT	GAGAAGGGTG	TGAAAAAAAT	GTTGAGTATG	AAAGAGTAAT	1020
GAATCGTATT	ATTAGAATGT	TAGGTGTAGA	TAAAGCAATT	CGTTATGTTA	TTTTGGTAA	1080
GATAATATCT	GTATTAACGG	GTTCCTGTT	AATAATGTTA	ATATCACACC	ATTTATCTAA	1140
AGACGCACAG	GGCTATTATT	ATACATTTAA	TTCAGTAGTG	GCACTACAGA	TAATATTTGA	1200
ATTGGGGCTA	TCAACGGTAA	TCATTCAATT	CGCTAGCCAT	GAAATGTCAG	CGTTAAAATA	1260
TGATTATTCT	GAACGAGATA	TTATAGGTGA	AAGTAAAAAT	AAGCAACGTT	ACCTATCGTT	1320
ATTCGGTTG	GCAATAAAAT	GGTATGCACT	AATAGCTTG	CTAATAATAT	TAATAGTCGG	1380
TCCCCATCGGG	TATGTTTTT	TTACGCAAAA	AGAAGGCTTA	GGTGTACCTT	GGCAAGGGC	1440
ATGGTTATTA	TTAACAAATAG	TTACAGCTTT	TAATATTTT	CTTGTTCCTG	TACTTTCTGT	1500
CGCTGAAGGG	AGTGGGTTAA	TTACTGATGT	GAATAAAATG	AGAATGTATC	AGTCGCTGTT	1560
AGCTGGTATA	TTGGCAGTAA	GCTTACTTAT	TAGTGGCTTT	GGACTATATG	CTACGCTCTGC	1620
AATAGCTATT	TCAGGGACTA	TCATATTCTC	CATATTTCA	TATAAGTATT	TTAAAAAAAT	1680
TTTCCTGCAA	TCTTAAAGC	ATAAAAATAA	ATATACTGAA	GGTGGTATTT	CATGGGTTAA	1740
TGAAATATTT	CCTATGCAAT	GGCGAATTGC	TCTAAGTTGG	ATGTCAGGGT	ATTTTATTTA	1800
TTTTGTTATG	ACCCCCATTG	CATTCAAATA	TTTCGGGCT	ATATATGCAG	GGCAGTTAGG	1860
GATGTCCTTA	ACATTATGCA	ATATGGTAAT	GGCTACGGGC	CTGGCTTGG	TATCCACTAA	1920
ATATCCAAA	TGGGGAGTAA	TGGTTCCAA	CAAACAGCTT	GCGGAACGTGA	GTAAATCGTT	1980
CAAAAGTGCA	GTAATGCAAT	CATCCTTTT	TGTCTTGACA	GGATTAACGT	GTGTATACAT	2040
TTCATTATGG	TTATTGAAAT	TATCTGGTTC	AAACATTGGC	GAGCGGTTTT	TGGGATTGCA	2100
GGATTTTTTC	TTTTTATCTT	TAGCAATTAT	TGGTAATCAC	ATTGTAGCTT	GCTTTGCAAC	2160
CTATATAAGA	GCGCATAAAA	CTGAAAAAAAT	GACATTGGCA	TCATGTATAA	TGGCTCTCTT	2220
GACTATAACT	ACAATGTTGT	TTGTTGCATA	TTTAGAGTAC	TCGAGGTTCT	ACATGTTAAT	2280
GTATGCAGCA	CTAACGTGGT	TATTTTGT	TCCTCAAAC	TATATAATCT	TTAAAAGATT	2340
CAAGAGTTCT	TATGAGTAA	AAACCTCTTC	TTACTATTGC	TATTCCGACA	TATAACCGCT	2400
CTTCATGTTT	GGCTCGTTA	CTTGATAGTA	TAATTCAACA	GGAGAACTAT	TGTCATGATG	2460
AACTCGAGGT	TATTGTTGT	GATAATGCTT	CAACAGATGA	AACAGCAAGA	ATAGCCAAGA	2520
GTGGCTTAGA	TAAAATAAGA	AATAGTACTT	ATCATCTAA	TGAAGAAAAC	TTAGGAATGG	2580
ATGGTAACTT	CCAGAAATGT	TTTGAGTTAT	CAAATGGAAA	ATATCTTG	ATGATTGGCG	2640
ATGATGATCT	AATAGTCAA	AATGGTATTT	CGAAGGTTTT	TTCGATATTA	AAGTCCCGGC	2700
CTGCATTAGA	TATGGTGTAT	GTAAATTCAAG	CAGCAAAGAC	TGAGTTAAC	TATAATGCTG	2760
ATGTGAGGAC	GTCATTCTAC	ACAAATGATG	TAGATTTAT	TTCAAGACGTG	AAAGTTATGT	2820

- 78 -

TCACGTTAT TTCTGGAATG ATATGTAAGA AACTGATGC AATTGTCAAA GCCGTTGGTA	2880
TTTCAGTCC GCAAACACT GGAAAATATC TTATGCATT AACATGGCAA TTGCCATTAC	2940
TTAACACAGGG TGGAGAGTTC GCAGTTATCC ATAATAATAT AATTGAGGCT GAGCCAGATA	3000
ATTCAGGTGG ATATCATTAA TATAAGGTTT TTTCTAATAA TCTTGCACAC ATCTTGATG	3060
TTTTTATCC CAGAGAGCAC CGTGTAAGTA AAAGAGTCG CGCATCAGCA TGTTTATTCT	3120
TACTTAACCT CATAGGCGAT GAAGATAAAA CCAAAAATTT TGCTACAAAT AATTATTTAA	3180
GAGATTGCGA TAGTGCATT ATAGATTAA TTATATATAA ATATGGGCTT AGGTTTTCT	3240
ATCTATATCC TAAAACGTG CCTTTATTAA GAAAAATAAA ATATATTATA AAGACGGTTT	3300
TAATGCGAA ATAAAAATTA TTCAAGATGG TTTGCTGAAAC ACGACTTATAA GGACTATCTA	3360
ATGTTTGCT ATAGTTAAG ATTAAAATTA AATCTTATCA TATCATTATT GAGTAAAGTT	3420
AGGCGGAAAT CAAAAGCAAAT GTTCTTGTT CTGCTTAGCG GATATGATTT TAAAATGGTT	3480
GGGAAGAATT TTAAATTGAA TGTCAAACCT TACTCTGCAA AAAATAACAC CTCTTCCAAA	3540
TGGGGTAGTA TGCGGGTTGG TGATAACTGC TGGATTGAAG CTGTATATAA TTATGGTGAT	3600
GAAAAATTG AACCTTATTT GTACATAGGT GATCGTATAT GTTTAAGTGA TAATGTTCAT	3660
ATTTCTTGCAT TATCATGTTT AATTTAGAA AACGATATAT TAATTGGTAG CAAAGTTTAT	3720
ATAGGCGATC ATAGCCATGG CAGTTATAAA GTATGCAGTC CGAAAATAGA ACCGCCAGCA	3780
AATAAGCCAT TAGGTGATAT TGCTCCTATT AAAATAGGTA ATTGCTGCTG GATTGGAGAT	3840
AATGCAGTAA TTCTGGCTGG TAGTGAAATT TGTGATGGCT GTGTAATCGC AGCTAATTCA	3900
GTCGTCAAGG ATTTAAAAGT CGATAAGCCA TGTTTAATTG GTGGGGTTCC TGCTAAAGTA	3960
ATAAAGGTAT TTAAATGAA ATGTTTTAT CAGTATTGT ATACCGTCTT ATAATAGAGC	4020
TGAGTTTTA GAGCCACTAC TGGATAGCAT ATATAATCAA GATTATTGTT TAAAGAATAA	4080
TGATTTGAG GTCATTGTTT GTGAAGATAA ATCTCCACAG AGAGATGAGA TAAAATCTAT	4140
TATCGAAAAC TATAAAGCAA AAAATAATAA ACAAAATCTT TATGTTAATT TCAATGAAGA	4200
TAATTAGGC TATGATAAGA ATTTAAAAAA ATGCATTAGT TTGACGACAG GTAAATATTG	4260
CATGATCATG GGCAACGATG ATCTATTAGC AGATGGAGCG TTATCAAAAA TAGTGAAAGT	4320
TTTGAAGGCT AATCCTGAAA TTGTATTGGC TACGCGAGCG TATGGTTGGT TTAAGGAAAA	4380
TCCGAATGAG TTATGTGATA CTGTTCGTCA TTTAACAGAC GATACTTTAT TTCAGCCGGG	4440
GGCTGATGCC ATTAAATTT TCTTCGTAG AGTTGGAGTT ATTCAGGCT TTATTGTC	4500
TGCTGAAAAA GCAAAAAAAC TATCGAGTGA TTTATTTGAT GGGCGTTAT ATTATCAAAT	4560
GTACCTTGCT GGTATGCTAA TGGCTGAAGG TCAGGGATAC TATTTAGCG ACGTGATGAC	4620
ATTGTCGAGG GATACAGAGG CTCCGTACTT TGGTAACGCT GGAACGTGAAA AAGGAGTTTT	4680
CACCCGGGG GGGTATAAAC CAGAGGGCCG TATACATATG GTTGAAGGCT TGTTGCTAAT	4740
TGCAAAATAT ATAGAAGATA CAACAAAAAT TGATGGCGTT TATGCTGGAA TTAGAAAAGA	4800
CTTAGCGAAC TATTTTATC CTTATATTAGC AGATCAACTC GACTTGCCTC TTTTAACTTA	4860

TATTAAAATG ATAAATAAAAT TTGGGAAAT GGGATTTCA AATGAAAAGC TTTTCTATGT	4920
GCATGCCTT TTAGGGTATG TACTAAAACG GAGGGGCTAT GATGCTTAA TTAAATACAT	4980
TCGTAGCAAA AAAGGCGGTA CTCCGCGTCT TGGTATTTAA CCTCCACTTT CAAAAAATGT	5040
TATGAATATA CTTCTTGCTG CGATATTAGG CGTTAACTTA TTTTCTCCAT ATATTAGTTC	5100
GTGGATGGTG GGTATGCTGC CATTTCACC AGGAGCAATC CTAAGGGATG TACTCAATGT	5160
ATTTTTGTG GCGTTAGTGC TAGTTCGATT TGTCAATTGAT AGGAAAAAAA CTTATTC	5220
GTTGGTTTT ACTATTTTT CATGGTCGGC GGTAATACTA TGGGTAATAG CGTTAACTAT	5280
ATTCTCACCG GATAAAATTC AAGCAATTAT GGGGGGGCGG AGTTATATTT TATTCCCGGC	5340
AGTTTCTATA GCATTAGTGA TTTTAAAAGT ATCATACCCG CAATCCTTAA ATATTGAAA	5400
AATAGTTGC TACATAATTT TTCTAATGTT TATGGTTGCG ACAATATCTA TTATTGATGT	5460
ACTAATGAAT GGAGAGTCA TTAAATTGCT CGGATATGAT GAGCATTATG CAGGAGAAC	5520
ATTAAACTTA ATTAATAGCT ATGATGGGAT GGTCCGGGCT ACAGGGGTT TTAGTGTGC	5580
TCTCAATTT GGATATATGC TCACATTAGG TGTTTGTG TGATGGAGT GTTTTCCC	5640
AGGATATAAA AGATTATTGA TGCTTATTAT TAGTTTGTG CTATTATAG CGATCTGCAT	5700
GAGTCTTACT AGAGGAGCAA TACTTGTGC TGCGCTTATT TACGCACTTT ATATAATTC	5760
AAATCGGAAG ATGCTTTTT GTGGAATAAC TTTATTTGTA ATAATTATAC CCGTTTTAGC	5820
AATTCTACT AATATTTTG ACAACTATAC AGAAATTTG ATCGGCAGGT TTACAGATTC	5880
GTCTCAGGCA TCGCGTGGAT CTACACAGGG GCGGATAGAT ATGGCAATTAA ATTCAATTAA	5940
CTTCCTGTCA GAACATCCAT CAGGTATAGG TCTGGGTACT CAAGGTTCA GAAACATGCT	6000
TTCGGTAAAA GATAATAGGT TAAATACGGA TAATTATTT TTCTGGATCG CCCTTGAGAC	6060
TGGTATTATT GGCTTAATCA TAAATATTAT TTATCTGGCA AGTCAATTAA ATTCTTCAAC	6120
TTTACTAAAT AGAATATATG GCAGTCATTG TAGCAATATG CACTATAGAT TATATTTCT	6180
CTTGGAAAGT ATATATTTA TAAGTCAGC GTTAAGTTCA GCACCTTCGT CATCAACTTT	6240
TTCTATATAT TATTGGACAG TTTAGCTTT GATTCCATTAA TAAATAGACG 6300	6360
GTGCACGCGA TAATGAATAA TAAAAAGGTT TTGATGGATA TTAGTTGGTC TAATAAAGGG	6420
GGGATTGGAC GTTTTACTGA TGAAATTTCT AAAACTACTAT GTGATATATC TAAGGAGGAA	6480
CTATATAGAA AATGTGCTTC TCCGCTGGCC CCATTAGGTT TAGCAGTCAT TATTTTCTG	6540
CGAAAGAAAA CTGATGTGGT TTTTCTTCCT GGCTATATTC CACCACTTT TTGTTGAA	6600
AAGTCATAA TAACAATACA TGATCTAAAT CATCTGGATT TAAATGATAA TTCTCTCTT	6660
TTTAAGAGGT TATTTATATAA TTTTATAATA AAGCGCGGTT GTAGAAAAGC ATATAAAATA	6720
TTTACAGTTT CGAATTTTC AAAAGAAAGA ATAGTAGCAT GGTCAGGTGT AAACCCTAAT	6780
AAAATAGTC CCGTATATAA TGGGGTATCT AGTCTATTTA ATGCCGATGT AAAACCATTG	6840
AATTTAGGCT ATAAATATTT GCTATGTGTA GGAAACAGAA AAACTCATAA GAATGAGAAG	6900
TGTGTTATAT CTGCCTTGC CAAAGCAGAT ATTGATCCAT CAATAAAACT CGTTTTACT	

- 80 -

GGTAATCCTT	GTAATGATTT	AGAAAAACTA	ATAATACAAC	ATGGTTAAG	TGAACGTGTA	6960
AAGTTCTTG	GGTCGTGTC	TGAAAAAGAT	TTACCATCGT	TATATAAGGG	CTCGTTAGGA	7020
TTAGTTTCC	CTTCTTTATA	TGAAGGTTT	GGATTACCTG	TAGTGGAGGG	CATGGCCTGT	7080
GGTATTCTG	TATTAACCTC	TCTAACTTCA	TCATTGCCAG	AGGTGGCTGG	AGATGCAGCG	7140
ATTCTTGTG	ACCCCTTTTC	GGAAGATGCT	ATTACTAAAG	GAATTCGAG	GTAAATTAAT	7200
GATTCTGAAC	TTCGTAAGCA	TTTAATCCAA	AAGGGGCTTT	TGCGGGCAAA	GAGGTTCAAT	7260
TGGCAAAACG	TGGTTAGTGA	GATTGAAATG	GTACTGACAG	AGGCATGTGA	TGAAATAAAA	7320
TGAAATAAAA	ATATCTCTCG	TTCATGAGTG	GTTATTAAGT	TATGCAGGCT	CCGAACAGGT	7380
ATCATCTGCC	ATCCTGCATG	TTTTCTGTA	AGCGAAGTTA	TATTCGGTGG	TTGATTTCT	7440
AACGGATGAA	CAAAGAAGAC	ATTTCTGGG	GAAATATGCG	ACTACCACAT	TTATTCAAAA	7500
TTTACCTAAA	GCTAAAAAAAT	TTTACCAAGAA	ATATTTACCA	CTAATGCCAC	TGGCTATTGA	7560
ACAACTTGAT	TTATCAGATG	CTAATATCAT	CATTAGTAGC	GCCCATTCCG	TTGCAAAAGG	7620
TGTTATTTCC	GGACCAGATC	AGCTTCACAT	TAGCTATGTT	CATTCTCCTA	TTCGATATGC	7680
GTGGGATTAA	CAGCATCAGT	ACCTTAATGA	GTCTAACCTG	AATAAAGGAA	TTAAAGGTTG	7740
GTTAGCAAAA	TGGCTTCTTC	ACAAAATACG	AATTTGGGAT	TCTCGAACCG	CAAATGGGTT	7800
TGATCATTTT	ATAGCTAATT	CTCAATATAT	CGCGCGTAGA	ATTAAGGAA	TATACAGACG	7860
TGAGGCTTCA	GTTATATATC	CGCCTGTAGA	TGTGGATAAT	TTGAAAGTAA	AAAATGAAAA	7920
GCAAGACTAT	TATTTCACAG	CATCCCGTAT	GGTACCCCTAC	AAACGTATTG	ATCTTATTGT	7980
CGAACGCTTT	AGTAAAATGC	CGGAAAAGAA	ATTAGTAGTT	ATTGGTGATG	GACCGGAGAT	8040
GAAAAAAATA	AAGAGCAAGG	CTACAGACAA	TATAAAATTG	CTCGGTTATC	AATCTTTCC	8100
TGTTTTAAAA	GAGTATATGC	AGAGGCCAG	GGCGTTGTT	TTTGCAGCGG	AAGAGGACTT	8160
TGGAATAATA	CCTGTCGAAG	CTCAAGCTTG	CGGTACCCCT	GTTATTGCCT	TTGGGAAGGG	8220
TGGGGCCTTA	GAAACCGTTC	GCCCACTAGG	TGTAGAGGAA	CCGACTGGCA	TTTTCTCAA	8280
GGAACAGAAT	ATTGCTTCTT	TGCATGAAGC	TGTTAGTGAA	TTGAAAAAA	ATGCATCATT	8340
TTTTACATCT	CAGGCTTGT	GAAAAAATGC	AGAAAATTT	TCTCGATCAA	GATTGAAACA	8400
AGAATTAAAG	AACTTTGTTA	ATGAAAAGTG	GAATCTTTC	AAAACAGAAC	AGATTATTAA	8460
ACGTTAATTA	TGGTTTATTG	AATGTCTAAA	TTAATACCA	TAATAATGGC	CGGTGGGATT	8520
GGTAGCCGTT	TGTGGCCACT	TTCACGTGAA	GAGCATCCGA	AAACAGTTTT	AAGCGTAGAT	8580
GGTGAATTAT	CTATGCTGCA	AAACACCATT	AAAAGATTGA	CTCCTTTTT	GGCTGGAGAA	8640
CCTTTAGTCA	TTTGTAAATGA	TAGTCACCGC	TTCCCTGTG	CTGAACAACT	TCGAGCTATA	8700
AATAAAACTAG	CAAATAACAT	CATATTAGAG	CCAGTGGGGC	GTAATACAGC	CCCAGCTATA	8760
GCGCTGGCCG	CTTTTGTTC	ACTTCAGAAT	GTCGTGATG	AAGACCCGCT	TTTGCTTGTC	8820
CTTGCTGCGG	ATCATGTCAT	CCCGGATGAG	AAAGTGTTC	TTAAAGCTAT	CAATCACGCT	8880
GAATTTTTTG	CAACACAAGG	TAAGCTAGTA	ACGTTGGTA	TTGTACCCAC	ACAGGCCGAA	8940

ACTGGCTACG	GTTATATTTG	TAGAGGTGAA	GCAATCGGGG	AAGATGCTTT	TTCTGTAGCC	9000
GAATTTGAG	AGAACGCTGA	TTTCGATACA	GCGCGTCATT	ATGTAGAATC	AGAGAAATAT	9060
TATTGAAACA	GCGGTATGTT	CCTATTCGT	GCAAGTAGTT	ACTTACAAGA	ATTAAAGGAT	9120
CTGTCCCCG	ATATTTACCA	AGCATGTGAA	AATGCGGTAG	GGAGTATTAA	TCCTGATCTT	9180
GATTTTATCC	GTATTGATAA	AGAACGATTC	GCAATGTGCC	CTAGTGATTC	TATCGATTAT	9240
GCGGTAATGG	AACATACTAG	GCATGCAGTT	GTCGTACCGA	TGAATGCCGG	CTGGTCAGAT	9300
GTGGGGTCAT	GGTCTTCACT	GTGGGATATT	TCTAAGAAAG	ATCCACAACG	TAATGTATTA	9360
CATGGCGATA	TTTTGCATA	TAATAGTAAA	GATAATTATA	TCTATTCTGA	AAAATCGTT	9420
ATTAGTACAA	TGGGAGTAAA	TAATTTAGTT	ATCGTGCAGA	CAGCAGATGC	ATTATTAGTA	9480
TCTGATAAAG	ATTCAGTCCA	GGATGTTAAA	AAAGTTGTTG	ATTATTAAA	AGCTAATAAT	9540
AGAAAACGAAAC	ATAAAAAAAC	TTTAGAGGTT	TTCCGACCGT	GGGGAAAATT	TAGCGTAATT	9600
CATAGTGGCG	ATAATTATTT	AGTTAAAAGA	ATAACTGTTA	AACCAGCGC	GAAGTTTGCT	9660
GCTCAGATGC	ATCTCCATCG	TGCTGAGCAT	TGGATAGTGG	TATCTGGTAC	TGCTTGTATT	9720
ACTAAGGGGG	AAGAAATTTT	TACAATTCG	GAGAATGAAT	CAACATTAT	ACCTGCTAAT	9780
ACAGTCATA	CGTTAAAAAA	CCCCGCGACT	ATTCCATTAG	AACTAATAGA	AATTCAATCT	9840
GGCACCTATC	TTGCGGAGGA	TGATATTATT	CGCCTGGAGA	AACATTCTGG	ATATCTGGAG	9900
TAATGAATTG	ATGAAAATA	TATATAATAC	TTACGATGTT	ATCAACAAAT	CTGGAATTAA	9960
TTTTGGAACC	AGTGGTGCCC	GCGGCCTTGT	TACCGATTTC	ACACCGAAG	TTTGCGCACG	10020
ATTTACCATT	TCCTTTTGA	CAGTAATGCA	GCAAAGATT	TCATTACAA	CGGTTGCGCT	10080
CGCAATTGAT	AATCGTCAA	GCAGTTACGC	GATGGCTCAA	GCTTGTGCCG	CTGCTTTGCA	10140
AGAAAAAGGA	ATTAAAACCG	TTTACTATGG	CGTAATTCCA	ACACCTGCTT	TAGCTCATCA	10200
ATCAATTCC	GATAAAGTAC	CTGCAATCAT	GGTTACTGGC	AGTCATATCC	CTTTTGACCG	10260
TAATGGCCTG	AAATTTATA	GACCAGATGG	TGAAATTACT	AAAGATGATG	AGAATGCTAT	10320
TATTCATGTT	GATGCCTCAT	TTATGCAGCC	TAAGCTTGAA	CAATTGACAA	TTTCCACAAT	10380
CGCTGCTAGA	AATTATATTC	TACGATATAC	CTCATTATTT	CCAATGCCAT	TCTTGAAAAA	10440
TAAGCGCATT	GGAATTATG	AGCATTCTAG	TGCGGGTCGT	GATCTCTATA	AGACGTTATT	10500
CAAATGTTG	GGTGCTACAG	TTGTTAGTTT	AGCAAGGAGC	GACGAATTG	TTCCATTGCA	10560
TACTGAAGCT	GTAAGTGAAG	ATGATAGAAA	TAAAGCAATC	ACATGGCAA	AAAAATATCA	10620
GTTAGATGCT	ATATTTCAA	CTGATGGTGA	TGGAGATCGC	CCTCTGATAG	CTGACGAATA	10680
TGGAAATTGG	TTAAGAGGGAG	ATATATTAGG	CCTTCTGTGC	TCTCTCGAAT	TAGCTGCTGA	10740
TGCAGTCGCT	ATTCCTGTA	GCTGCAACAG	TACAATCTCA	TCTGGTAACT	TTTTAAACA	10800
TGTGGAACGA	ACAAAGATTG	GTTCACCCCTA	TGTGATTGCA	GCATTGCTA	AATTATCTGC	10860
AAACTATAAT	TGTATAGCTG	GTGTTGAAGC	GAATGGTGGC	TTTCTGCTAG	GTAGCGATGT	10920
TTATATTAAT	CAGCGTTTAC	TTAAGGCATT	ACCAACACGT	GATGCTTTAT	TACCTGCCAT	10980

- 82 -

TATGCTTCTG	TTGGTAGCA	AGGACAAAAG	TATTAGTGAG	CTTGTAAAAA	AACTTCCTGC	11040
TCGCTATACC	TATCAAACA	GATTACAGGA	TATAAGTGT	AAAACAAGTA	TGTCTTTAAT	11100
AAATCTTGGT	CTGACAGATC	AAGAGGATT	TTTGCAGTAT	ATTGGTTTA	ATAAACATCA	11160
TATATTACAT	TCTGATGTTA	CTGATGGCTT	TAGAATCACT	ATCGATAACA	ACAATATTAT	11220
TCATTTACGA	CCTTCAGGCA	ATGCCCTGA	GTTCGTTGC	TATGCGGAGG	CTGACTCGCA	11280
AGAGGATGCA	TGTAATATTG	TTGAAACTGT	TCTCTCTAAT	ATCAAAAGCA	AACTGGGTAG	11340
AGCTTAATGC	TGTTGATAAT	AGAGCGTTTC	TTTCCAGTAA	TACTTGTCT	GGTTATCTGG	11400
TACCCAAGTT	GAGGGTGAGA	ATTAAATGGA	TCGTTTTGAT	AATAAGTATA	ACCCAAATT	11460
ATGCAAAATA	TTATTGGCTA	TATCAGATTT	ACTGTTTTTT	AATGTAGCCT	TATGGGCATC	11520
GTTAGGAGTT	GTATATTAA	TCTTGATGA	AGTTCAGCGA	TTTGTACAC	AAGAGCAATT	11580
AGATAATCGA	TTTATATCAC	ATTTTATTCT	ATCTATAGTA	TGC GTTGGAT	GGTTTTGGGT	11640
TCGACTGCGT	CACTATACAT	ATCGAAAGCC	ATTCTGGTAT	GAGTTGAAAG	AGGTTATTG	11700
TACTATCGTT	ATTTTGCTG	TGTTTGATTT	GGCTTTAATT	GC GTTTACAA	AATGGCAGTT	11760
TTCACGCTAT	GTCTGGGTGT	TTTGTGGAC	TTTGCCATA	ATCCTGGTGC	CTTTTTTCG	11820
CGCACTTACA	AAGCATTAT	TGAACAAGCT	AGGTATCTGG	AAGAAAAAAA	CTATCATCCT	11880
TGGGAGCGGA	CAGAATGCTC	GTGGTGCATA	TTCTGCGCTG	CAAAGTGAGG	AGATGATGGG	11940
GT TTGATGTT	ATCGCTTTT	TTGATACGGA	TGC GT CAGAT	GCTGAAATAA	ATATGTTGCC	12000
GGTGATAAAG	GACACTGAGA	CTATTGGGA	TTTAAATCGT	ACAGGTGATG	TCCATTATAT	12060
CCTTGCTTAT	GAATACACCG	AGTTGGAGAA	AACACATT	TGGCTACGTG	AACTTTCAA	12120
ACATCATTGT	CGTTCTGTTA	CTGTCGTCCC	CTCGTTAGA	GGATTGCCAT	TATATAATAC	12180
TGATATGTCT	TTTATCTTA	GCCATGAAGT	TATGTTATTA	AGGATACAAA	ATAACTTGGC	12240
TAAAAGGTG	TCCCGTTTC	TCAAACGGAC	ATTTGATATT	GT TTGTTCAA	TAATGATTCT	12300
TATAATTGCA	TCACCACTTA	TGATTATCT	GTGGTATAAA	GTTACTCGAG	ATGGTGGTCC	12360
GGCTATTTAT	GGTCACCAGC	GAGTAGGTG	GCATGGAAAA	CTTTTCCAT	GCTACAAATT	12420
TCGTTCTATG	GTTATGAATT	C				12441

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22080 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iv) ANTI-SENSE: YES
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: *S. enterica* serovar *typhimurium* (serogroup B)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GAATTCTGGGA	GGCGCAATGA	AAGTCAGCTT	TTTCTGCTG	AAATTTCCAC	TCTCATCGGA	60
AACCTTTGTG	CTGAATCAGA	TTACTGCGTT	TATTGATATG	GGCCATGAGG	TGGAGATTGT	120
CGCGTTACAA	AAAGGCATA	CCCAACATAC	TCACGCCGCC	TGGGAGAAGT	ATGGCCTGGC	180
GGCGAAAACC	CGCTGGTTAC	AGGATGAGCC	CCAGGGACGG	CTGGCGAAAC	TGCGCTACCG	240
GGCATGTAAA	ACGCTGCCGG	GGCTGCATCG	GGCGCGACC	TGGAAAGCGC	TCAATTTCAC	300
CCGCTATGGC	GATGAATCAC	GCAATTGAT	CCTTCCGCG	ATTTGCGCGC	AGGTGAGCCA	360
GCCTTTGTG	GCGGATGTGT	TTATCGCACA	CTTGGTCCG	GCAGGGCGTGA	CGGCGGCCAA	420
ACTACCGCAA	CTGGCGTGC	TTCGCGCAA	AATCGCAGT	ATTTCCACG	GGATTGATAT	480
CTCTAGTCGT	GAGGTGCTCA	GTCATTACAC	GCCGGAGTAT	CAGCAGTTGT	TTCGTCGTGG	540
CGATCTGATG	CTGCCCATCA	GCGATCTGTG	GGCCGGTCGC	CTGAAAAGTA	TGGGCTGTCC	600
GCCGGAAAAG	ATTGCCGTTT	CGCGCATGGG	CGTCGACATG	ACGCGTTTA	CCCATCGTTC	660
GGTAAAAGCG	CCAGGGATGC	CGCTGGAGAT	GATTTCGTC	GCAGCGCTGA	CAGAAAAAAA	720
AGGCCTGCAT	GTGGCGATTG	AAGCCTGTCG	GCAACTGAAA	GCACAGGGCG	TGGCGTTTCG	780
CTACCGCATT	CTGGGGATTG	GCCC GTGGGA	ACGTCGGCTG	CGCACGCTCA	TCGAGCAGTA	840
TCAGCTAGAG	GATGTCATTG	AGATGCCGGG	GTTTAAACCG	AGCCATGAAG	TGAAGGCGAT	900
GCTGGATGAC	GCCGATGTTT	TTTGCTGCC	GTCGATTACC	GGTACGGATG	GCGATATGGA	960
AGGTATTCCG	GTAGCGCTGA	TGGAGGCGAT	GGCGGTAGGG	ATTCCCGTGG	TATCTACCGT	1020
GCATAGCGGT	ATTCCGGAAC	TGGTGGAGGC	CGGCAAATCC	GGCTGGCTGG	TGCCGGAAA	1080
CGATGCGCAG	GGCCTGGCGG	CCCGACTCGC	TGAGTTCAGC	CGGATTGACC	ACGACACGCT	1140
GGAGTCGGTG	ATCACGCGCG	CCCGTGGAAA	AGTGGCGCAA	GATTTAACATC	AGCAGGGCGAT	1200
TAATCGCCAG	TTAGGCCAGC	TGCTACAAAC	GATATAAACG	AGGTGGTATG	CCCGCGACTA	1260
AATTCTCCCG	ACGTACCCCTC	CTGACGGCAG	GTTCTGCGCT	TGCTGTTCTT	CCTTTCTGC	1320
GCGCCTTGCC	GGTACAGGGCG	CGTGAACCTC	GCGAGACCGT	CGATATTAAG	GATTATCCGG	1380
CGGATGACGG	TATCGCTCG	TTCAAACAGG	CCTTCGCCGA	CGGACAGACC	GTGGTCGTAC	1440
CGCCAGGATG	GGTGTGTGAA	AATATCAATG	CGGCGATAAC	GATTCCGGCG	GGAAAAACGC	1500
TGCGGGTACA	GGCGCGGTG	CGTGGGAATG	GCCGGGGACG	GTTTATTTTG	CAGGACGGGT	1560
GTCAGGTGGT	GGGGGAGCAG	GGCGGCAGTC	TGCACAATGT	GACGCTGGAT	GTTCGCGGGT	1620
CGGACTGTGT	GATTAAAGGC	GTGGCGATGA	GGGGCTTTGG	CCCCGTGCGC	CAAATTTCA	1680
TCGGTGGTAA	GGAACCGCAG	GTGATGCGTA	ATCTCATTAT	CGATGACATC	ACCGTTACCC	1740
ACGCCAACTA	CGCCATTCTC	CGCCAGGGAT	TTCATAACCA	AATGGATGGC	GCGCGGATTA	1800
CGCATAGCCG	CTTAGCGAT	TTACAGGGGG	ACGCCATTGA	GTGGAATGTC	GCGATTACAG	1860
ACCGCGACAT	CCTGATTTC	GATCATGTCA	TCGAACGCAT	TAATTGTACC	AATGGCAAAA	1920

TCAACTGGGG GATCGGCATC	GGGCTGGCGG GTAGCACCTA	TGACAACAGT TATCCTGAAG	1980
ACCAGGCAGT	AAAAAACTTT GTGGTGGCCA	ATATTACCGG ATCTGATTGC	2040
TGCACGTAGA	AAATGGCAAA CATTTCGTCA	TTCGCAATGT CAAAGCCAAA	2100
CCGGTTTCAG	TAAAAATGCG GGTATTGATA	ACGCAACGAT CGCAATTAT	2160
ATTTCGTCAT	TGATAATATT GATATGACGA	ATAGTGCCGG GATGCTCATC	2220
TCGTTAAAGG	AAAATACCTG TCAATTCCGC	AAAACCTTAA ATTAAACGCT	2280
ATAATCGCCA	GGTTGCTTAT AAATTACCGG	GCATTCAAAT TTCCCTCCGGC	2340
CTTTGTGCGC	CATCACCAAT GTACGGATGA	CGCGTGCTAC GCTGGAACTG	2400
CGCAGCACCT	CTTCTGCGC AATATCAACG	TGATGCAAAC TTCAGCGATT	2460
TAAAAATGCA	TTTCGATTTG CGTAAAGATG	TACGTGGTCA ATTTATGGCC	2520
CGCTGCTTTC	CCTCGCTAAT GTTCATGCCA	TCAATGAAAA CGGGCAGAGT	2580
TCGACAGGAT	TAATCACCAA ACCGTGAATG	TCGAAGCAGT GAATTTTCG	2640
GGGGAGGGTA	AGTACCGCTA TTTTACGAA	AATTCTGGG AAAAAGTTGT	2700
TGTTATGGTG	CCGACTAAGA CGTAATGTAG	AGCGTGCCAT CATTATCCCT	2760
TAATTCATGC	TGGCGAAAAC AAGCTAAAGA	GCTATAATTC AGCAACCATT	2820
AAGAAACAAT	GATGAATTTG AAAGCAGTTA	TACCGTAGC GGGTTTGGGT	2880
TGCCTGCCAC	CAAGGCAATC CCAAAAGAGA	TGCTACCGAT CGTCGACAAG	2940
AGTACATTGT	CGATGAGATT GTGGCTGCAG	GGATCAAAGA AATCGTGCTG	3000
CGTCTAAAAA	CGCCGTTGAG AACCACTTCG	ACACCTCTTA TGAACTTGAA	3060
AGCAGCGCGT	TAAGCGTCAG CTTTGGCGG	AAGTGAATC TATCTGCCA	3120
CGATTATGAA	CGTCGCCAG GCGCAGCCGT	TAGGGCTGGG GCATTCTATT	3180
GTCCGGTCGT	GGGCATAAAC CCTTCATTG	TGGTACTCCC GGATATTATT	3240
CTACCGCCGA	TCCGCTGCGC TATAACCTTG	GGCGATGGT GGCGCGTTTC	3300
GTCGCAGCCA	GGTGCTGGCG AAGCGCATGA	AAGGTGATTG ATCGGAGTAT	3360
AGACGAAAGA	ACCTCTGGAT AATGAAGGCA	AAGTCAGCCG GATTGTGGAG	3420
AACCGGATCA	GCCGCAGACG CTGGATTCCG	TTTGATGGC GGTAGGCCGT	3480
CAGCCGACAT	CTGGCGGGAA CTGGAAAGAA	CGAACCGGG CGCCTGGGGC	3540
TCACCGATGC	CATTGCTGAA CTGGCGAAA	AACAGTCGGT TGACCGATG	3600
GTGACAGCTA	TGACTGCGGT AAAAAATGG	GCTACATGCA GGCATTG	3660
TGCGCAACCT	GAAAGAAGGA GCCAAGTCC	GTAAGAGCAT AGAGCAGCTT	3720
AAGTATTAAC	AACCGTGATA AATGGTTGGT	GATAAACATA ATAACGGCAG	3780
AAGCGGCAAG	TTGGCTGAAA CGAGTGTGA	CTGCCGTTTT AGTTTGTAT	3840
AGTAACAAGG	GGTTATCTGG AGCATTAA	TGCTGATTAA TCCTTGTTC	3900
CGGATGCAAT	TAATAAGACA ATTAGCGTTT	AAGTTTTAGT GAGCTTGCC	3960

AGGTTTGCAA CAAGTCGATA TGTACGCAGT GCACTGGTAG CTGATGAGCC AGGGGCGGTA	4020
GCGTGTGAA CGACTTGAGC AATTAATTT TATTGGCAA TTAAATACCA CATTAAATAC	4080
GCCTTATGGA ATAGAAAAGT GAAGATACTT ATTACTGGCG GGGCAGGTTT TATTGGATCA	4140
GCTGTTGTC GCCATATTAT TAAGAATACA CAGGACACTG TAGTTAATAT TGATAAATTA	4200
ACCTACGCCG GTAATCTTGA ATCCCTTCT GATATTCTG AAAGTAATCG CTACAATTTC	4260
GAACACGCGG ATATTTGTGA TTCCGCTGAA ATAACCGTA TTTTGAGCA GTACCAGCCG	4320
GACGCGGTGA TGCATTTGGC TCGGGAAAGT CATGTGGACC GTTCGATTAC CGGGCCAGCA	4380
GCATTTATTG AAACCAATAT CGTCGGCACC TATGCACTTC TTGAAGTTGC GCGTAAATAC	4440
TGGTCTGCCC TTGGCGAAGA TAAAAAAAAT AATTTTCGTT TTCACTCATAT TTCCACTGAT	4500
GAAGTTTACG GCGATTTACC GCATCCTGAT GAAGTTGAAA ACAGCGTTAC GCTGCCGTTA	4560
TTTACTGAAA CGACGGCATA TGCGCCAAGT AGCCCCTATT CTGCGTCAA AGCATCCAGC	4620
GATCATTAG TCCGTGCCTG GCGCGTACCA TATGGTCTAC CAACGATCGT TACCAATTGT	4680
TCTAATAACT ATGGCCCTTA TCACCTCCCT GAAAAACTGA TTCCGTTGGT CATTGAAAC	4740
GCACTGGAAG GAAAGCCTT GCCAATTAT GGCAGGGGG ATCAGATTG CGATTGGCTA	4800
TATGTAGAAG ATCATGCTCG CGCGCTTCAT ATGGTAGTGA CTGAAGGCAA GGCAGGGGAG	4860
ACTTATAACA TTGGTGGACA CAATGAGAAG AAAATCTCG ATGTGGTATT TACCATCTGT	4920
GATCTGCTGG ATGAGATTGT ACCCAAAGCG ACTTCTTATC GTGAACAAAT CACTTATGTC	4980
GCGGATCGTC CGGGCCATGA TCGTCGTTAT GCCATTGATG CAGGTAAAAT TAGCCGCGAA	5040
TTAGGCTGGA AACCGCTGGA GACCTTGAA AGCGGTATTC GTAAAACAGT GGAATGGTAC	5100
CTTGCAAATA CTCATGGGT AAACAATGTT AAAAGTGGGG CGTATCAGAG TTGGATAGAA	5160
CAGAACTATG AAGGACGCCA GTAATGAATA TCTTACTTTT TGGTAAGACA GGGCAAGTAG	5220
GCTGGGAGTT GCAACGTTCT CTGGCACCGG TAGGGAATCT GATTGCCCTG GATGTCCATT	5280
CAAAAGAGTT TTGCGGTGAT TTTAGTAATC CGAAAGGCGT TGCCGAAACC GTTCGTAAGC	5340
TTCGTCCGGA TGTGATTGTT AACGCAGCAG CCCATACTGC AGTAGATAAA GCAGAGTCTG	5400
AACCAGAACT GGCAGGTTA CTTAACGCCA CCAGTGTGGA AGCCATCGCT AAAGCAGCCA	5460
ACGAAACTGG CGCATGGGT GTGCATTATT CAACCGATTA TGTATTCCT GGTACCGCG	5520
ATATCCCATG GCAGGAAACG GACGCTACGT CGCCGCTGAA TGTCTATGGC AAAACCAAAC	5580
TGGCGGGAGA AAAGGCCCTG CAGGATAACT GCCCTAAACA CCTTATCTTC CGCACCAAGT	5640
GGGTTTATGC AGGTAAGGGC AATAATTCG CAAAGACAAT GCTTCGCTG GCGAAAGAGC	5700
GTCAGACACT TTCAGTCATT AACGATCAGT ACGGTGCGCC AACCGGTGCG GAATTACTGG	5760
CTGACTGTAC GGCGCATGCG ATCCGTGTGG CGTTAAATAA ACCAGAAAGTC GCAGGTCTT	5820
ACCATCTGGT TGCCGGGGGA ACCACAAACCT GGCATGACTA CGCGGCCCTA GTCTTGACG	5880
AGGCGCGCAA AGCAGGGATA ACGCTTGCGC TGACTGAGCT TAATGCTGTG CCGACCAGCG	5940
CCTACCCGAC GCGGGCGAGC AGACCAGGCA ATTGCGGTCT CAATACTGAA AAGTTTCAGC	6000

GTAATTTGA CCTTATTCTG CCTCAATGGG AATTAGGAGT TAAGCGTATG CTGACTGAAA	6060
TGTTTACGAC GACAACCATC TAATAAATTT AAATGCCAT CAGGGCATT TCTATGAATG	6120
AGAAATGGAA ATGAAAACGC GTAAGGGCAT TATTTAGCG GGGGGCTCCG GCACCCGTCT	6180
TTATCCGGTG ACCATGGCGG TAAGTAAGCA ATTGCTACCA ATTTATGATA AACCGATGAT	6240
TTACTATCCC CTTTCCACGC TTATGCTGGC AGGCATTGG AGATATCTGA TCATCAGTAC	6300
GCCACAGGAC ACGCCGCGTT TTCAACAACT GCTGGGAGAC GGCAGCCAGT GGGGGCTGAA	6360
TCTTCAATAT AAAGTACAGC CAAGCCCGGA TGGCTTAGCA CAGGGCTTTA TTATTGGTGA	6420
AGAGTTCATT GGTACATGATG ATTGTGCATT AGTGCTGGGT GACAATATCT TCTATGGTCA	6480
TGATTTACCA AAGTTAATGG AAGCTGCCGT TAATAAAGAA AGTGGTGCTA CCGTCTTCGC	6540
TTATCATGTA AACGATCCGG AGCGCTACGG TGTGGTTGAG TTTGACCAAA AGGGCACAGC	6600
CGTTAGTCTG GAAGAAAAAC CATTACAACC GAAGAGTAAT TACGCGTAA CGGGGCTGTA	6660
TTTTTATGAT AATAGCGTGG TGGAGATGGC GAAAAATCTT AAGCCTTCCG CTCGCGGTGA	6720
GTTAGAAATC ACGGATATTA ACCGTATCTA TATGGAGCAG GGAAGATTGT CTGTCGCTAT	6780
GATGGGGCGC GGTTATGCCT GGCTGGATAC AGGGACGCAT CAGAGTTGTA TAGAGGCCAG	6840
TAATTTTATT GCAACCATCG AAGAACGCCA GGGGCTAAAAA GTGTCCTGCC CGGAAGAGAT	6900
CGCATTTCGT AAAAATTAA TAAATGCACA ACAGGTTATA GAACTGGCCG GCCCATTATC	6960
AAAAAATGAT TATGGCAAAT ATTTGCTGAA GATGGTAAA GGTTATAAG TGATGATTGT	7020
GATTAAAAACA GCAATACCAAG ATGTCTTGAT CTTAGAGCCT AAAGTTTTG GCGATGAGAG	7080
GGGATTCTTT TTTGAAAGTT ATAACCAGCA GACCTTGAA GAGTTGATTG GACGTAAAGT	7140
TACATTGTT CAAGATAATC ATTCAAAATC CAAAAAGAAC GTACTCAGAG GGCTACATT	7200
TCAGAGAGGA GAAAATGCAC AGGGGAAGTT AGTTCGTTGT GCTGTCGGTG AGGTTTTG	7260
TGTTGCGGTC GATATCCGAA AAGAATCGCC TACTTTGGT CAATGGGTTG GTGTAATCT	7320
GTCTGCTGAG AATAAGCGAC AGCTTGGAT TCCAGAAGGT TTTGCTCATG GTTTGTTAC	7380
TCTTAGTGAG TATGCAGAGT TTCTGTACAA AGCAACTAAT TATTACTCAC CTTCATCGGA	7440
AGGTAGCATT CTATGGAATG ATGAGGCAAT AGGTATTGAA TGGCCTTTT CTCAGCTGCC	7500
TGAGCTTCA GCAAAAGATG CTGCAGCACC TTTACTGGAT CAAGCCTGT TAACAGAGTA	7560
AGCATCGTGT CTCATATTAT TAAGATTTT CCATCAAATA TTGAATTTC CGGTAGAGAG	7620
GATGAATCAA TCCTCGATGC TGCGCTATCG GCTGGTATCC ATCTGAACA TAGCTGCAA	7680
GCGGGTGATT GTGGTATCTG TGAGTCCGAT TTGTTGGCGG GAGAAGTTGT TGACTCCAA	7740
GGTAATATTT TTGGACAGGG TGATAAAATA CTAACCTGCT GCTGAAACC TAAAACGCC	7800
CTTGAGCTAA ATGCGCATT TTTCTCGAA CTAGCTGGAC AGACAAAAAA AATTGTCCCA	7860
TGCAAGGTAA ATAGTGTGT ACTGGTTCA GGCGATGTTA TGACTTGAA GTTACGCACA	7920
CCACCAACAG CAAAAATTGG CTTCCCTCCA GGGCAGTATA TCAATTACAA TTATAAAGGT	7980
GTAACTCGCA GTTATTCTAT CGCTAATAGT GATGAGTCGA ATGGTATTGA GTTGCATGTA	8040

- 87 -

AGGAATGTTCA CCAATGGTCA GATGAGTTCG CTCATTTTG GGGAGTTACA AGAAAATACT	8100
CTTATGCGCA TTGAAGGGCC TTGCGGAACA TTTTTTATTG GTGAAAGTGA CAGACCTATA	8160
ATCTTCCTTG CAGGCGGTAC TGGATTGCT CCAGTTAAAT CAATGGTTGA GCATCTCATT	8220
CAGGGAAAAT GTCGTCGTGA GATCTACATT TACTGGGAA TGCAATATAG TAAAGATTT	8280
TACTCTGCAT TACCGCAGCA GTGGAGTGAA CAGCACGACA ACGTTCATTA TATCCCTGTT	8340
GTTTCTGGTG ATGACGCCGA ATGGGGGGGA AGAAAGGGAT TTGTCCATCA TGCCGTGATG	8400
GATGATTGGT ATTCTCTAGA GTTCTTCGAT ATATATGCAT GTGGTTCAACC TGTGATGATC	8460
GATGCCAGTA AAAAGGACTT TATGATGAAA AATCTCTTG TAGAACATTT CTATTCTGAT	8520
GCATTTACCG CACTAATAAA TATTGAGGAT AATTATGAA AGCGGTCACTC CTGGCTGGTG	8580
GACTTGGTAC CAGACTAAGT GAAGAAACAA TTGTAAAACC AAAACCGATG GTAGAAATTG	8640
GTGGCAAGCC TATTCTTG G CACATTATGA AAATGTATTG TGTGCATGGT ATCAAGGATT	8700
TTATTATCTG CTGTGGTTAT AAAGGATATG TGATTAAAGA ATATTTGCG AACTACTTCC	8760
TTCACATGTC AGATGTAACA TTCCATATGG CTGAAAACCG TATGGAAGTT CACCATAAAC	8820
GTGTTGAACC ATGGAATGTC ACATTGGTTG ATACGGGTGA TTCTTCAATG ACTGGTGGTC	8880
GTCTGAAACG TGGTGTGAA TACGTAAAAG ATGACGAGGC TTTCTGTTT ACTTATGGTG	8940
ATGGCGTTGC CGACCTTGAT ATCAAAGCGA CTATCGATT CCATAAGGCT CACGGTAAGA	9000
AAGCGACTTT AACAGCTACT TTTCCACCAG GACGCTTG CGCATTAGAT ATCCGAGCTG	9060
GTCAGGTCCG GTCATTCCAG GAAAAACCGA AAGGCGATGG GGCAATGATC AATGGTGGTT	9120
TCTTTGTGTT GAATCCATCG GTTATCGATC TCATCGATAA CGATGCAACA ACCTGGGAAC	9180
AAGAGCCATT AATGACATTG GCACAACAGG GGGAGTTAAT GGCTTTGAA CACCCAGGTT	9240
TCTGGCAGCC GATGGATACC CTACGTGATA AAGTTTACCT CGAAGGGCTG TGGGAAAAAG	9300
GTAAAGCTCC GTGGAAAACC TGGGAGTAAC TAGATGATTG ATAAAAATTT TTGGCAAGGT	9360
AAACGTGTAT TCGTTACCGG CCATACTGGC TTTAAAGGAA GCTGGCTTTC GCTATGGCTG	9420
ACTGAAATGG GTGCAATTGT AAAAGGCTAT GCACTTGATG CGCCAACTGT TCCAAGTTA	9480
TTTGAGATAG TGGTCTTAA TGATCTTATG GAATCTCATA TTGGCGACAT TCGTGATTT	9540
GAAAAGCTGC GCAATTCTAT TGCAGAATTG AAGCCAGAAA TTGTTTCCA TATGGCAGCC	9600
CAGCCTTACG TGCGCCTATC TTATGAACAG CCAATCGAAA CATACTCAAC AAATGTTATG	9660
GGTACTGTCC ATTTGCTTGA AACAGTTAAG CAAGTAGGTA ACATAAAGGC AGTCGTAAAT	9720
ATCACCAAGTG ATAAGTGCTA CGACAATCGT GAGTGGGTGT GGGCTATCG TGAGAACGAA	9780
CCCATGGGAG GGTACGATCC ATACTCTAAT AGTAAAGGTT GTGCAGAATT AGTCGGTCT	9840
GCATTCCGGA ACTCATTCTT CAATCCTGCA AATTATGAGC AACATGGCGT TGGTTGGCG	9900
TCTGTGAGGG CTGGTAATGT CATAAGGCGGA GGCATTGGG CTAAGAACCG TTTAATTCCC	9960
GATATTCTGC GCTCATTGAA AATAAACAG CAGGTTATTA TTCGAAACCC ATATTCTATC	10020
CGTCCCTGGC AGCATGTAAGT GGAGCCTCTT TCTGGTTACA TTGTGGTGGC GCAACGCTTA	10080

TATACAGAAG	GTGCTAAGTT	TTCTGAAGGA	TGGAATTCG	GCCCGCGTGA	TGAAGATGCC	10140
AAGACGGTCG	AATTATTATTGT	TGACAAGATG	GTCACGCTT	GGGGTGATGA	TGCAAGCTGG	10200
TTACTGGATG	GTGAGAATCA	TCCTCATGAG	GCACATTACC	TGAAACTGGA	TTGCTCTAAA	10260
GCAAATATGC	AATTAGGATG	GCATCCGCGT	TGGGGATTGA	CTGAAACACT	TGGTCGCATC	10320
GTAAAATGGC	ATAAAAGCATG	GATTCGCGGC	GAAGATATGT	TGATTTGTT	AAAGCGTGAA	10380
ATCAGCGACT	ATATGTCTGC	AACTACTCGT	TAAGAAAATA	AGTTTAAGGA	ATCAAAGTAA	10440
TGACAGCAAA	TAACCTGCGT	GAGCAAATCT	CTCAGCTTGT	CGCTCAGTAT	GCGAATGAGG	10500
CATTGAGCCC	GAAACCTTTT	GTTGCAGGTA	CAAGCGTTGT	GCCTCCTTCC	GGGAAGGTTA	10560
TTGGTGCCTAA	AGAGTTACAA	TTGATGGTTG	AGGCGTCTCT	TGATGGATGG	CTAACTACTG	10620
GTCGTTCAA	TGATGCCTTT	GAAAAAAAAC	TTGGGGATT	TATTGGGGTT	CCTCATGTTT	10680
TAACGACAAAC	ATCTGGCTCT	TCGGCAAAC	TGCTGGCACT	GACTGCGCTG	ACTTCCCCAA	10740
AATTAGGCGA	GCGAGCTCTC	AAACCTGGTG	ATGAGGTTAT	TACTGTCGCT	GCTGGCTTCC	10800
CGACTACAGT	TAACCCGGCG	ATCCAGAATG	GTAAATACC	GGTATTGCG	GATGTTGATA	10860
TCCCGACATA	TAATATCGAT	GCCTCTCTCA	TTGAAGCTGC	AGTTACTGAG	AAATCAAAAG	10920
CGATAATGAT	CGCTCATACA	CTCGGTAATG	CATTTAACCT	GAGTGAAGTT	CGTCGGATTG	10980
CCGATAAAATA	TAACTTATGG	TTGATTGAAG	ACTGCTGTGA	TGCCCTTGGG	ACGACTTATG	11040
AAGGCCAGAT	GGTAGGTACC	TTTGGTGACA	TCGGAACCGT	TAGTTTTAT	CCGGCTCACCC	11100
ATATCACAAT	GGGTGAAGGC	GGTGTGTAT	TCACCAAGTC	AGGTGAACTG	AAGAAAATTA	11160
TTGAGTCGTT	CCGTGACTGG	GGCCGGGATT	GTTATTGTGC	GCCAGGATGC	GATAACACCT	11220
GCGGTAAACG	TTTGGTCAG	CAATTGGGAT	CACTTCCTCA	AGGCTATGAT	CACAAATATA	11280
CTTATTCCCA	CCTCGGATAT	AATCTAAAA	TCACGGACAT	GCAGGCAGCA	TGTGGCTGG	11340
CTCAGTTGGA	GCGCGTAGAA	GAGTTGTAG	AGCAGCGTAA	AGCTAACTTT	TCCTATCTGA	11400
AACAGGGCTT	GCAATCTTGC	ACTGAATTCC	TCGAATTACC	AGAAGCAACA	GAGAAATCAG	11460
ATCCATCCTG	TTTGGCTTC	CCTATCACCC	TGAAAGAAC	TAGCGGTGTT	AACC GTGTCG	11520
AACTGGTGA	ATTCTTGTAT	GAAGCAAAA	TCGGTACACG	TTTACTGTT	GCTGGAAATC	11580
TGATTGCCA	ACCGTATTTT	GCTAATGTGA	AATATCGTGT	AGTGGGTGAG	TTGACAAATA	11640
CCGACCGTAT	AATGAATCAA	ACGTTCTGGA	TTGGTATTAA	TCCAGGCTTG	ACTACAGAGC	11700
ATTTAGATTA	TGTAGTTAGC	AAGTTGAAG	AGTTCTTGG	TTTGAATTTC	TAATTCAATT	11760
TATTCTATCT	GGTGATTGCG	ATGACCTTTT	TGAAAGAATA	TGTAATTGTC	AGTGGGGCTT	11820
CCGGCTTAT	TGGTAAGCAT	TTACTCGAAG	CGCTAAAAAA	ATCGGGGATT	TCAGTTGTCG	11880
CAATCACTCG	AGATGTAATA	AAAAATAATA	GTAATGCATT	AGCTAATGTT	AGATGGTGCA	11940
GTTGGGATAA	TATCGAATTAA	TTAGTCGAGG	AGTTATCAAT	TGATTCTGCA	TTAATTGGTA	12000
TCATTCACTT	GGCAACAGAA	TATGGCATA	AAACATCATC	TCTCATAAAAT	ATTGAAGATG	12060
CAAATGTTAT	AAAACCATTA	AAGCTTCTTG	ATTTGGCAAT	AAAATATCGG	GC GGATATCT	12120

- 89 -

TTTTAAATAC AGATAGTTT TTTGCCAAGA AAGATTTAA TTATCAACAT ATGCGGCCTT	12180
ATATAATTAC TAAAAGACAC TTTGATGAAA TTGGGCATTA TTATGCTAAT ATGCATGACA	12240
TTTCATTGT AAACATGCGA TTAGAGCATG TATATGGGCC TGGGGATGGT GAAAATAAAT	12300
TTATTCCATA CATTATCGAC TGCTTAAATA AAAAACAGAG TTGCGTAAA TGTACAACAG	12360
GCGAACAGAT AAGAGACTT ATTTTGTAG ATGATGTGGT AAATGCTTAT TTAACTATAT	12420
TAGAAAATAG AAAAGAAGTA CCTTCATATA CTGAGTATCA AGTTGGAACG GGTGCTGGGG	12480
TAAGTTGAA AGATTTCTG GTTTATTGCA AAAATACTAT GATGCCAGGT TCATCGAGTA	12540
TATTTGAATT TGGTGCATA GAGCAAAGAG ATAATGAAAT AATGTTCTCT GTAGCAAATA	12600
ATAAAAATTT AAAAGCAATG GGCTGGAAAC CAAATTCGA TTATAAAAAA GGAATTGAAG	12660
AACTACTGAA ACGGTTATGA GATTTTCATG ATCTTTAAT AAATAAATCG TTAACAAATT	12720
AGTCGCGTTA TGTTGTAAAA ACTAAGTCGT TTAATTGCAT AGTGAAGTT CAATTGTTAA	12780
AAATTCCGAG TCATTTAATT GTTGCAGGTT CATCATGGTT ATCCAAAATA ATAATTGCCG	12840
GGGTGCAGTT AGCAAGTATT TCATATCTT TTTCTATGCT AGGTGAAGAG AAATATGCAA	12900
TCTTTAGTTT GTTAACTGGT TTATTAGTAT GGTGTAGCGC TGTTGATTTT GGCATAGGTA	12960
CAGGACTGCA AAATTATATA TCAGAATGCA GAGCCAAAAA CAAAAGTTAT GATGCATATA	13020
TTAAATCAGC ATTACATCTA AGCTTTATAG CTATTATTTT TTTTATTGCT TTATTTATA	13080
TTTTTCTGG GGTAAATTCC GCTAAATATC TTTCTTCTT TCATGAGGTA TTACAGGACA	13140
AAACCAGAAAT GCTCTTTTT ACCTCATGTC TGGTTTCAG TTCTATTGGA ATCGGAGCTA	13200
TTGCTTATAA AATACTTTT GCCGAATTGG TCGGGTGGAA AGCTAATCTA TTAAACGCAT	13260
TATCTTATAT GATAGGTATG CTCGGCTTGC TATATATATA CTATAGGGGG ATCTCAGTTG	13320
ACATAAAATT ATCACTAATA GTCCTGTATC TTCCAGTGGG TATGATTCA TTGTGCTATA	13380
TTGTATATAG ATACATAAAAG CTTTATCATG TTAAAACAAC AAAATCTCAT TATATAGCAA	13440
TTTACGTAG ATCTTCAGGG TTTTTCTTT TTACTTTATT ATCGATAGTG GTGCTTCAA	13500
CAGATTATAT GGTCATTCT CAAAGGCTAA CTCCTGCTGA TATTGTTCAA TATACAGTAA	13560
CGATGAAAAT TTTGGTTTA GTCTTTTTA TTTATACTGC TATTGTTCAA GCATTATGGC	13620
CTATATGTGC TGAATTGAGA GTCAAACAGC AATGGAAAAA ACTTAACAAA ATGATAGGTG	13680
TCAATATTTT GCTGGCTCA CTATATGTTG TTGGATGTAC AATATTTATT TATTTATTAA	13740
AAGAACAGAT ATTTTCAGTA ATAGCCAAAG ATATTAATTA TCAAGTTCT ATTTTATCTT	13800
TTATGTTAAT TGGCATATAT TTCTGTATTC GCGTTGGTG TGACACTTAT GCAATGTTAT	13860
TGCAAAGTAT GAATTATTTA AAAATACTTT GGATATTAGT ACCACTACAA GCAATAATTG	13920
GTGGAATAGC ACAATGGTAT TTTCTAGTA CGCTTGGAAAT CAGTGGAGTG CTGCTTGGCT	13980
TGATTATATC TTTTGCTTTA ACTGTTTTT GGGGGCTTCC ACTAACTTAC TTAATTAAGG	14040
CAAATAAGGG ATAATCATAT GCTTATATCA TTTTGTATTC CAACTTATAA TAGAAAACAA	14100
TATCTTGAAG AGTTGTTGAA TAGTATAAAT AATCAGGAAA AATTTAATT AGATATTGAG	14160

- 90 -

ATATGTATAT CAGATAATGC CTCTACTGAT GGTACAGAGG AAATGATTGA TGTTTGGAGG	14220
AACAATTATA ATTTCCCAAT AATATATCGG CGTAATAGCG TTAACCTTGG GCCAGATAGG	14280
AATTTCTTG CTTCACTATC CCTTGCGAAT GGGGATTATT GTTGGATATT TGGCAGTGAT	14340
GATGCTCTTG CGAAAGACTC GTTAGCGATA TTACAAACTT ATCTCGATTC TCAAGCAGAT	14400
ATATATTTAT GTGACAGAAA AGAGACCGGG TGTGATTTAG TTGAGATTAG AAACCTCAT	14460
CGTTCTTGGC TCAGAACAGA TGATGAACCTT TATGTGTTA ATAATAATTT AGATAGGGAA	14520
ATCTATCTCA GTAGATGCTT ATCTATTGGT GGTGTATTTA GCTATCTAAG TTCTTTAATA	14580
GTAAAAAAAG AACGATGGGA TGCCATTGAT TTTGATGCGT CCTATATTGG CACTTCCTAT	14640
CCTCATGTAT TTATCATGAT GAGCGTATTT AATACGCCAG GGTGCCTTT GCATTATATA	14700
TCAAAACAC TCGTAATATG CCGAGGAGAT AATGATAGTT TCGAGAAGAA AGGAAAGGCC	14760
AGACGAATTT TAATTGATTT TATTGCATAT TTAAAATTAG CTAATGATTT TTACAGTAA	14820
ATATATCTT TAAAACGAGC ATTTGAAAAT GTTTGCTAA AAGAGAGACC ATGGTTATAT	14880
ACAACTTGG CTATGGCATG TTATGGCAAT AGTGATGAAA AAAGAGATTT ATCTGAATTT	14940
TATGCAAAGC TAGGTTGTA TAAAATATG ATCAACACTG TACTTCGATT TGGGAAACTA	15000
GCATATGCAG TGAAAAATAT TACCGTGCTT AAGAATTAA CTAAACGGAT ATTAAGTAG	15060
TAGTAAGTTA TTATATTGAG ATTAATGTA GATTAACTT TTCTGGATTC AGCTAGATTT	15120
ACGTTACTGA CTTTCTTTT TAATGAAAAT CATATTTGAT ATATATAAAT AAATTTGGAT	15180
AGCTTAACTA CTTAGATGTT TTTTCTGGG AATGTTAGTA TAATAATATA TTTCTTTATG	15240
ATTGTTTTG TAGTGTAACTA CTGCCGTAT TACATTAACCT CTATTATTA GAATTACACC	15300
TAGTGTAAGC TTCGTAATAT TATTTATCCT TATGATTATT GCTTAAAGA TGCGTATGGA	15360
AAAACGGAGA GCTATTCAAT GATCGAAAC CTATCACGTT TAGTAAAAG TGGTACGGGA	15420
ATGTGGCAAT ACTCGATTAA ATTTTAACG GCACTGCGAG AAATAGCTGA TGTTGACGCA	15480
ATAATCTGTA GCAAGGTACA CGCTGATTAT TTTGAAAAGC TCGTTATGC AGTAGTTACT	15540
GTTCCGAATA TTGTTAGCAA CACATCAAAA ACATCGCGAC TTAGACCATT AGTATGGTAT	15600
GTATATAGTT ACTGGCTTGC GCTGAGGGTT TTAATTAAGT TTGGTAATAA AAAATTGGTG	15660
TGTACTACAC ATCACACTAT CCCCTTACTG AGAAACCAAA CGATAACCGT ACATGATATA	15720
AGACCTTTT ATTATCCAGA TAGTTTATT CAGAAAGTGT ATTTTCGCTT TTTATTAAAA	15780
ATGTCGTTA AGCGATGTA GCATGTAAACCTT ACGGTATCTT ATACCGTTAA AGATAGCATT	15840
GCTAAAACCTT ATAATGTAGA TAGTGAGAAA ATATCAGTAA TTTATAATAG TGTTAATAAA	15900
TCTGTTTTA TACAAAAAAA AGAAAAAGAG AATTACTTT TAGCTGTTGG TGCAAGTTGG	15960
CCACATTTAATTCATTC ATTCATTTAATAA AATAAAAAAG TTTGGTCTGA CTCTTATAAT	16020
TTAATTATTG TATGTGGTCG TACTGACTAT GCAATGTCTC TCCAACAAAT GGTCGTTGAT	16080
CTGGAACCTAA AAGATAAAAGT GACTTTTTA CATGAAGTCT CATTAAATGA ATTAAAGATT	16140
TTATATTCTA AAGCCTACGC GCTTGTAACTT CCATCTATTG ATGAGGGTTT TGGTATACCT	16200

CCTATTGAAG CGATGGCATC AAATACTCCA GTTATAGTGT CCGATATACC AGTATTCAT	16260
GAAGTGTAA CCAATGGTGC ATTATATGTG AATCCGGATG ATGAAAAAAG CTGGCAGAGT	16320
GCAATTAAA ATATAGAGCA GTTGCCTGAT GCAATTCCC GATTTAACAA CTATGTCGCA	16380
CGGTATGACT TTGATAATAT GAAGCAGATG GTTGGCAATT GGTTGGCGGA ATCAAAATAA	16440
ATGAAAATAA CATTAAATTAT TCCCACATAT AATGCAGGGT CGCTTGGCC TAATGTTCTG	16500
GATGCGATTA AGCAGCAAC TATATATCCG GATAAATTGA TTGTTATAGA CTCAGGTTCT	16560
AAAGATGAAA CGGTTCCGTT AGCCTCAGAC CTGAAAAATA TATCAATATT TAATATTGAC	16620
TCTAAAGATT TTAATCATGG AGGAACCAGA AATTTAGCAG TTGCAAAAC TCTGGACGCT	16680
GATGTTATAA TTTTCTAAC GCAAGATGCA ATTCTCGCGG ATTCGGATGC AATTAAAAAT	16740
TTGGTTATT ATTTTCAGA TCCATTGATA GCAGCGTTT GTGGTAGACA ACTTCCTCAT	16800
AAAGATGCTA ATCCCCTTGC AGTGCATGCC AGAAATTAA ATTATAGTTC AAAATCTATT	16860
GTTAAAAGTA AGGCAGATAT AGAAAAATTG GGTATTAAAA CTGTATTAT GTCCAATTCT	16920
TTTGCTGCCT ATCGCCGTT CGTTTGAA GAGTTAAGTG GGTTCCCTGA ACATACAATT	16980
CTTGGCAGG ATATGTTAT GGCGGCTAAG ATGATTCAAG CGGGTTATAA GTTCGCCTAC	17040
TGCGCTGAAG CGGTGGTAAG ACACCTCCAT AATTATACCC CGCGAGAAGA GTTCAACGA	17100
TATTTGATA CTGGTGTATT TCATGCTTGT TCTCCGTGGA TTCAGCGTGA CTTTGGCGGA	17160
GCCGGTGGTG AGGGTTCCG CTCGTAAAA TCAGAGATTG AATTCTGCT TAAAAATGCA	17220
CCGTTCTGGA TTCCAAGAGC TTTATTAACA ACCTTGCTA AATTCTGGG TTACAAATT	17280
GGCAAGCATT GGCAATCTT ACCGTTGTCT ACATGTCGCT ATTTAGCAT GTACAAGAGT	17340
TATTGGAATA ATATCCAATA TTCTCGTCA AAAGAGATAA AATAATGTC TTTCTTCCC	17400
GTAATTATGG CTGGCGGCAC AGGTAGCCGT TTATGGCCGC TTTCACCGA ATATCATCCG	17460
AAGCAGTTTC TAAGCGTTGA AGGTAAACTA TCAATGCTGC AAAATACTAT AAAGCGATTA	17520
GCTTCACCTT CTACAGAAGA ACCCGTTGTC ATTTGCAATG ACAGACACCG TTTCTTAGTC	17580
GCTGAACAAC TCCGTGAAAT TGACAAGTTA GCAAATAATA TTATTCTCGA ACCGGTAGGC	17640
CGTAATACTG CACCAAGCGAT CGCTCTTGCC GCGTTTGTC CGCTCCAGAA TGCTGATAAT	17700
GCTGATCCTC TTTGTTGGT TCTTGCTGCA GATCATGTGA TTCAGGATGA AATAGCTTT	17760
ACGAAAGCTG TCAGACATGC TGAAGAATAC GCTGCAAATG GTAAGCTTGT AACCTTTGGT	17820
ATTGTTCCAA CGCATGCTGA AACGGTTAT GGATATATTC GTCGTGGTGA GTTGATAGGA	17880
AATGACGCTT ATGCAGTGGC TGAATTGTC GAGAAACCGG ATATCGATAC CGCCGGTGAC	17940
TATTTCAAAT CAGGGAAATA TTACTGGAAT AGCGGTATGT TTTTATTCG TGCAAGCTCT	18000
TATTTAAACG AATTAAAGTA TTTATCACCT GAAATTATA AAGCTTGTGA AAAGGCGGTA	18060
GGACATATAA ATCCCGATCT TGATTTATT CGTATTGATA AAGAAGAGTT TATGTCATGC	18120
CCGAGTGATT CTATCGATTA TGCAGTTATG GAGCACACAC AGCATGCGGT GGTGATACCA	18180
ATGAGCGCTG GCTGGTGGGA TGTGGTTCC TGGTCCTCAC TTTGGGATAT ATCGAATAAA	18240

GATCATCAGA GAAATGTTT AAAAGGAGAT ATTTTCGCAC ATGCTTGTAA TGATAATTAC	18300
ATTTATTCCG AAGATATGTT TATAAGTGC GATTGGTGTAA GCAATCTGT CATTGTTCAA	18360
ACAAACAGACG CTTTACTGGT GGCTAATAAA GATACAGTAC AAGATGTTAA AAAAATTGTC	18420
GATTATTAA AACGGAATGA TAGGAACGAA TATAAACAAAC ATCAAGAAGT TTTCCGCC	18480
TGGGGAAAAT ATAATGTGAT TGATAGCGC AAAAATTACC TCGTTCGATG TATCACTGTT	18540
AAGCCGGGTG AGAAATTGT GGCGCAGATG CATCACCACC GGGCTGAGCA TTGGATAGTA	18600
TTATCCGGGA CTGCTCGTGT TACAAAGGGA GAGCAGACTT ATATGGTTTC TGAAAATGAA	18660
TCAACATTAA TTCCTCCGAA TACTATTCA GCGCTGGAAA ATCCTGGAAT GACCCCCCTG	18720
AAGTTAATTG AGATTCAATC AGGTACCTAT CTTGGTGAGG ATGATATTAT TCGTTTAGAA	18780
CAACGTTCTG GATTTTCGAA GGAGTGGACT AATGAACGTA GTTAATAATA GCCGTGATGT	18840
TATTTATTCA TCAGGTATTG TGTTTGGAAC GAGTGGGGCT CGCGGTCTTG TAAAAGATT	18900
TACACCTCAG GTATGTGCTG CTTTACGGT TTCATTGTT GCCGTTATGC AGGAACATT	18960
TTCCTTGAT ACCGTAGCAT TGGCAATAGA TAATCGTCCA AGTAGTTATG GGATGGCTCA	19020
GGCGTGTGCT GCTGCATTGG CGGATAAAGG CGTTAACTGT ATTTTTATG GAGTGGTACC	19080
AACCCCAGCT TTGGCCTTTC AGTCTATGTC TGACAATATG CCTGCGATAA TGGTTACGG	19140
AAGTCATATT CCATTGAGC GGAACGGCCT CAAGTTTAT CGTCCTGATG GTGAAATCAC	19200
GAAACATGAT GAGGCTGCGA TCCTTAGTGT TGAAGATACG TGCAGCCATT TAGAGCTTAA	19260
AGAAACTCATA GTTTCAGAAA TGGCTGCTGT TAATTATATA TCTCGTTATA CATCTTATT	19320
TTCTACTCCA TTCTGAAAA ATAAGCGTAT TGGTATTTAC GAACATTCAA GCGCTGGCG	19380
TGATCTTAT AAGCCTTTAT TTATTGCATT GGGGGCTGAA GTCGTTAGCT TGGGTAGAAG	19440
CGATAATTT GTACCTATAG ATACAGAGGC TGTAAGCAA GAGGATCGGG AAAAAGCTCG	19500
CTCATGGGCT AAAGAGTTCG ATTTAGATGC CATATTCTCG ACAGATGGGG ATGGTGATCG	19560
CCCTCTTATT GCTGATGAGG CCGGTGAGTG GCTAAGAGGC GATATACTAG GTCTATTATG	19620
TTCACTTGCA TTGGATGCGA AAGCCGTCGC TATTCTGTT AGTTGTAACA GCATAATT	19680
TTCTGGCCGC TTTTTAAAC ATGTTAAGCT TACAAAAATT GGCTCGCCT ATGTTATCGA	19740
AGCTTTAAT GAATTATCGC GGAGTTATAG TCGTATTGTC GGTTTGAAG CCAATGGCG	19800
TTTTTATTA GGAAGCGACA TCTGTATTAA CGAGCAGAAT CTTCATGCCT TACCAACTCG	19860
TGATGCTGTA TTACCAAGCAA TAATGCTGCT TTACAAAAGT AGGAATACCA GCATTAGCGC	19920
TTTAGTCAAT GAACTCCCAA CTCGTTACAC CCATTCTGAC AGATTACAGG GGATTACAAC	19980
TGATAAAAGT CAATCCTTAA TTAGTATGGG CAGAGAAAAT CTGAGCAACC TCTTAAGCTA	20040
TATTGGTTG GAGAATGAAG GTGCAATTTC TACAGATATG ACAGATGGTA TGCGAATTAC	20100
TTTACGTGAT GGATGTATTG TGCATTTGCG CGCTTCTGGT AATGCACCTG AGTTACGCTG	20160
CTATGCAGAA GCTAATTAT TAAATAGGGC TCAGGATCTT GTAAATACAA CGCTTGCTAA	20220
TATTAAAAAA CGATGCTTGC TGTAAAAAA TTGAATGTTA TTTACTTAAT ATGCCTATTT	20280

- 93 -

TATTTACATT ATGCACGGTC AGAGGGTGAG GATTAAATGG ATAATATTGA TAATAAGTAT	20340
AATCCACAGC TATGTAAAAT TTTTTGGCT ATATCGGATT TGATTTTTT TAATTTAGCC	20400
TTATGGTTTT CATTAGGATG TGTCTATTT ATTGTTGATC AAGTACAGCG ATTTATTCCCT	20460
CAAGACCAAT TAGATACAAG AGTTATTACG CATTTTATTT TGTCAGTAGT ATGTGTCGGT	20520
TGGTTTGGGA TTCGTTGCG ACATTATACT ATCCGCAAGC CATTGTTGGTA TGAGTTAAAA	20580
GAAATTTTC GTACGATCGT TATTTTGCT ATATTTGATT TGGCTCTGAT AGCCTTACA	20640
AAATGGCAGT TTTCACGCTA TGTCTGGGTG TTTGTTGGGA CTTTGCCCT AATCCTGGTG	20700
CCTTTTTTC GCGCACTTAC AAAGCATTAA TTGAACAAGC TAGGTATCTG GAAGAAAAAA	20760
ACTATCATCC TGGGGAGCGG ACAGAATGCT CGTGGTGCAT ATTCTGCGCT GCAAAGTGAG	20820
GAGATGATGG GGTTTGATGT TATCGCTTTT TTTGATAACGG ATGCGTCAGA TGCTGAAATA	20880
AATATGTTGC CGGTGATAAA GGATACTGAG ATTATTTGGG ATTTAAATCG TACAGGTGAT	20940
GTCCATTATA TCCTTGCTTA TGAATACACC GAGTTGGAGA AAACACATT TTGGCTACGT	21000
GAACTTCAA AACATCATTG TCGTTCTGTT ACTGTAGTCC CCTCGTTAG AGGATTGCCA	21060
TTATATAATA CTGATATGTC TTTTATCTTT AGCCATGAAG TTATGTTATT AAGGATACAA	21120
AATAACTTGG CTAAAAGGTC GTCCCGTTTT CTCAAACGGA CATTGATAT TGTTTGTCA	21180
ATAATGATTC TTATAATTGC ATCACCACTT ATGATTTATC TGTGGTATAA AGTTACTCGA	21240
GATGGTGGTC CGGCTATTAA TGGTCACCAAG CGAGTAGGTC GGCATGGAAA ACTTTTCCA	21300
TGCTACAAAT TTGTTCTAT GGTTATGAAT TCTCAAGAGG TACTAAAAGA ACTTTTGGCT	21360
AACGATCCTA TTGCCAGGGC TGAATGGAG AAAGATTAA AACTGAAAAA TGATCCTCGA	21420
ATCACAGCTG TAGGTCGATT TATACGTAAA ACTAGCCTTG ATGAGTTGCC ACAACTTTT	21480
AATGTACTAA AAGGTGATAT GAGCCTGGTT GGACCACGAC CTATCGTTTC GGATGAACTG	21540
GAGCGTTATT GTGATGATGT TGATTATTAT TTGATGGCAA AGCCGGGCAT GACAGGTCTA	21600
TGGCAAGTGA GTGGCGTAA TGATGTTGAT TATGACACTC GTGTTATT TGATTCCCTGG	21660
TATGTTAAA ACTGGACGCT TTGGAATGAT ATTGCCATTC TGTTAAAAC AGCGAAAGTT	21720
GTTTTGCAGC GAGATGGTGC GTATTAAGCT TACCGAGAAAG TACTGAATAA TAATTGTATA	21780
AATTAGCCTG CGTAAAATCT GAACGCATCA ATCGCTACCT TAATATCATA CCTTTGAGTT	21840
AACATACTAT TCACCTTAA CCTGCCATGA CCGTTGTGG CAGGGTTCC ACACCTGACA	21900
GGAGTATGTA ATGTCCAAGC AACAGATCGG CGTCGTCGGT ATGGCAGTGA TGGGGCGCAA	21960
CCTCGCGCTC AACATCGAAA GCCGTGGTTA TACCGTCTCC GTTTCAACC GCTCCCGTGA	22020
AAAGACCGAA GAAGTGATTG CCGAGAATCC CGGAAAAAG CTGGTGCCTT ATTACACGGT	22080

THE CLAIMS:

1. A nucleic acid molecule derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit, including a *wzx* gene or a *wzy* gene, or a gene with a similar function; the gene being involved in the synthesis of a particular bacterial polysaccharide antigen, wherein the sequence of the nucleic acid molecule is specific to the particular bacterial polysaccharide antigen.

2. A nucleic acid molecule derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit such as a *wzx* or *wzy* gene; the gene being involved in the synthesis of a particular bacterial O antigen, wherein the sequence of the nucleic acid molecule is specific to the particular bacterial O antigen.

3. A nucleic acid molecule derived from: a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit such as a *wzx* or *wzy* gene; the gene being involved in the synthesis of an O antigen expressed by *E. coli*, wherein the sequence of the nucleic acid molecule is specific to the O antigen.

4. A nucleic acid molecule derived from a gene encoding a transferase; or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit such as a *wzx* or *wzy* gene; the gene being involved in the synthesis of an O antigen expressed by *S. enterica*, wherein the sequence of the nucleic acid molecule is specific to the O antigen.

5. A nucleic acid molecule according to any one of claims 1 to 4 wherein the nucleic acid molecule is

approximately 10 to 20 nucleotides in length.

6. A nucleic acid molecule derived from a gene, the gene being selected from a group consisting of the
5 following sequences:

nucleotide position 739 to 1932 of SEQ ID NO:1;
nucleotide position 8646 to 9911 of SEQ ID NO:1;
nucleotide position 9901 to 10953 of SEQ ID NO:1;
nucleotide position 11821 to 12945 of SEQ ID NO:1;
10 nucleotide position 79 to 861 of SEQ ID NO:2;
nucleotide position 858 to 2042 of SEQ ID NO:2;
nucleotide position 2011 to 2757 of SEQ ID NO:2;
nucleotide position 2744 to 4135 of SEQ ID NO:2;
nucleotide position 5257 to 6471 of SEQ ID NO:2; and
15 nucleotide position 13156 to 13821 of SEQ ID NO:2;
which nucleic acid molecule is capable of hybridizing to
complementary sequence from said gene.

7. A nucleic acid molecule which is any one of
20 the oligonucleotides in Table 5 or 5A, with respect to the
genes *wbdH*, *wzx*, *wzy* and *wbdM*.

8. A nucleic acid molecule which is any one of
the oligonucleotides in Table 6 or 6A.

25 9. A nucleic acid molecule derived from a gene, the gene being selected from a group consisting of the
following sequences:

nucleotide position 1019 to 2359 of SEQ ID NO:3;
30 nucleotide position 2352 to 3314 of SEQ ID NO:3;
nucleotide position 3361 to 3875 of SEQ ID NO:3;
nucleotide position 3977 to 5020 of SEQ ID NO:3;
nucleotide position 5114 to 6313 of SEQ ID NO:3;
nucleotide position 6313 to 7323 of SEQ ID NO:3;
35 nucleotide position 7310 to 8467 of SEQ ID NO:3;
nucleotide position 12762 to 14054 of SEQ ID NO:4; and
nucleotide position 14059 to 15060 of SEQ ID NO:4;
which nucleic acid molecule is capable of hybridizing to

complementary sequences from said gene.

10. A nucleic acid molecule which is any one of the oligonucleotides in Table 7.

5

11. A nucleic acid molecule which is any one of the oligonucleotides in Table 8 with respect to the genes *wzx* and *wbaV*.

10 12. A method of testing a sample for the presence of one or more bacterial polysaccharide antigens, the method comprising the following steps:

15 (a) contacting the sample with at least one oligonucleotide molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing of oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the bacterial polysaccharide antigen; under 20 conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to at least one such gene of any bacteria expressing the bacterial polysaccharide antigen present in the sample and (b) detecting any specifically hybridised oligonucleotide 25 molecules.

13. The method according to claim 12, the method further comprising contacting the sample with a further at least one oligonucleotide molecule capable of specifically 30 hybridising to at least one sugar pathway gene under conditions suitable to permit the further at least one oligonucleotide molecule to specifically hybridise to at least one such sugar pathway gene of any bacteria expressing the bacterial polysaccharide antigen present in 35 the sample and detecting any specifically hybridised oligonucleotide molecules.

14. A method of testing a sample for the presence

of one or more bacterial polysaccharide antigens, the method comprising the following steps:

(a) contacting the sample with at least one pair of oligonucleotide molecules, with at least one 5 oligonucleotide molecule of the pair capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing of oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene; wherein 10 the gene is involved in the synthesis of the bacterial polysaccharide antigen; under conditions suitable to permit the at least one oligonucleotide molecule of the pair of molecules to specifically hybridise to at least such gene of any bacteria expressing the bacterial 15 polysaccharide antigen present in the sample and (b) detecting any specifically hybridised oligonucleotide molecules.

15. The method according to claim 14, the method 20 further comprising contacting the sample with a further at least one pair of oligonucleotide molecules, with at least one oligonucleotide molecule of the pair capable of specifically hybridising to at least one sugar pathway 25 gene under conditions suitable to permit the further at least one oligonucleotide molecule of the pair to specifically hybridise to at least one such sugar pathway gene of any bacteria expressing the bacterial polysaccharide antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules.

30

16. A method of testing a sample for the presence of one or more bacterial O antigens, the method comprising the following steps:

(a) contacting the sample with at least one 35 oligonucleotide molecule capable of specifically hybridising to: (i) a gene encoding an O antigen transferase, or (ii) a gene encoding an enzyme for transport or processing of the oligosaccharide or

polysaccharide units, including a *wzx* or *wzy* gene; wherein said gene is involved in the synthesis of the bacterial O antigen; under conditions suitable to permit the at least one oligonucleotide molecule to specifically hybridise to

5 at least one such gene of any bacteria expressing the bacterial O antigen present in the sample and

(b) detecting any specifically hybridised oligonucleotide molecules.

10 17. The method according to claim 16, the method further comprising contacting the sample with a further at least one oligonucleotide molecule capable of specifically hybridising to at least one sugar pathway gene under conditions suitable to permit the further at least one oligonucleotide molecule to specifically hybridise to at least one such sugar pathway gene of any bacteria expressing the bacterial O antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules.

15

20 18. The method according to claim 16 or 17 wherein the O antigen is expressed by *E. coli* or *S. enterica*.

25 19. The method according to claim 18 wherein the *E. coli* express the 0157 O antigen serotype or the 0111 O antigen serotype.

20 20. The method according to claim 18 wherein the *S. enterica* express the C2 or B O antigen serotype.

30 21. The method according to any one of claims 16 to 20 wherein the specifically hybridised oligonucleotide molecules are detected by Southern blot analysis.

35 22. A method of testing a sample for the presence of one or more bacterial O antigens, the method comprising the following steps:

(a) contacting the sample with at least one pair of oligonucleotide molecules, with at least one oligonucleotide molecule of the pair being capable of specifically hybridising to: (i) a gene encoding an O antigen transferase, or (ii) a gene encoding an enzyme for transport or processing of oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene; wherein the gene is involved in the synthesis of the bacterial O antigen; under conditions suitable to permit the at least one oligonucleotide molecule of the pair of molecules to specifically hybridise to at least one such gene of any bacteria expressing the bacterial O antigen present in the sample and

(b) detecting any specifically hybridised oligonucleotide molecules.

23. The method according to claim 22, the method further comprising contacting the sample with a further at least one pair of oligonucleotide molecules, with at least one oligonucleotide molecule of the pair capable of specifically hybridising to at least one sugar pathway gene under conditions suitable to permit the further at least one oligonucleotide molecule of the pair to specifically hybridise to at least one such sugar pathway gene of any bacteria expressing the bacterial O antigen present in the sample and detecting any specifically hybridised oligonucleotide molecules.

24. The method according to claim 22 or 23 wherein the O antigen is expressed by *E. coli* or *S. enterica*.

25. The method according to claim 24 wherein the *E. coli* are 0111 or the 0157 O antigen serotype.

35 26. The method according to claim 24 wherein the *S. enterica* express the C2 or B O antigen serotype.

27. The method according to any one of claims 22 to 26 wherein the method is performed according to the polymerase chain reaction method.

5 28. The method according to any one of claims 22 to 26 wherein the oligonucleotide molecules are selected from the group of nucleic acid molecules according to any one of claims 5 to 11.

10 29. A method for testing a food derived sample for the presence of one or more particular bacterial O antigens, the method being according to any one of claims 16 to 28.

15 30. A method for testing a faecal derived sample for the presence of one or more particular bacterial O antigens, the method being according to any one of claims 16 to 28.

20 31. A method for testing a sample derived from a patient for the presence of one or more particular bacterial O antigens, the method being according to any one of claims 16 to 28.

25 32. A kit comprising a first vial containing a first nucleic acid molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene, wherein said gene is involved in the synthesis of a bacterial polysaccharide.

30 33. The kit according to claim 32 further comprising in the first vial, or in a second vial, a second nucleic acid molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene, wherein

said gene is involved in the synthesis of a bacterial polysaccharide, and wherein the sequence of the second nucleic acid molecule is different from the sequence of the first nucleic acid molecule.

5 34. The kit according to claim 33 further comprising a nucleic acid molecule derived from a sugar pathway gene.

10 35. A kit according to claim 32 further comprising in the first vial, or in a second vial, a second nucleic acid molecule capable of specifically hybridising to a sugar pathway gene.

15 36. A kit according to any one of claims 32 to 35 wherein the nucleic acid molecules are approximately 10 to 20 nucleotides in length.

20 37. A kit comprising a first vial containing a first nucleic acid molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene, wherein said gene is involved in the synthesis of a bacterial O antigen.

25 38. The kit according to claim 37, further comprising in the first vial, or in a second vial, a second nucleic acid molecule capable of specifically hybridising to: (i) a gene encoding a transferase, or (ii) a gene encoding an enzyme for transport or processing oligosaccharide or polysaccharide units, including a *wzx* or *wzy* gene, wherein said gene is involved in the synthesis of a bacterial O antigen, and wherein the sequence of the second nucleic acid molecule is different from the sequence of the first nucleic acid molecule.

35 39. A kit according to claim 37 further comprising in the first vial, or in a second vial, a second nucleic acid molecule capable of specifically hybridising to a

sugar pathway gene.

40. The kit according to claim 38 further comprising a nucleic acid molecule derived from a sugar pathway gene.

5

41. The kit according to any one of claims 37 to 40 wherein the nucleic acid molecules are approximately 10 to 20 nucleotides in length.

10

42. The kit according to any one of claims 31 to 34 wherein the first and second nucleic acid molecules are according to any one of claims 5 to 11.

1/58

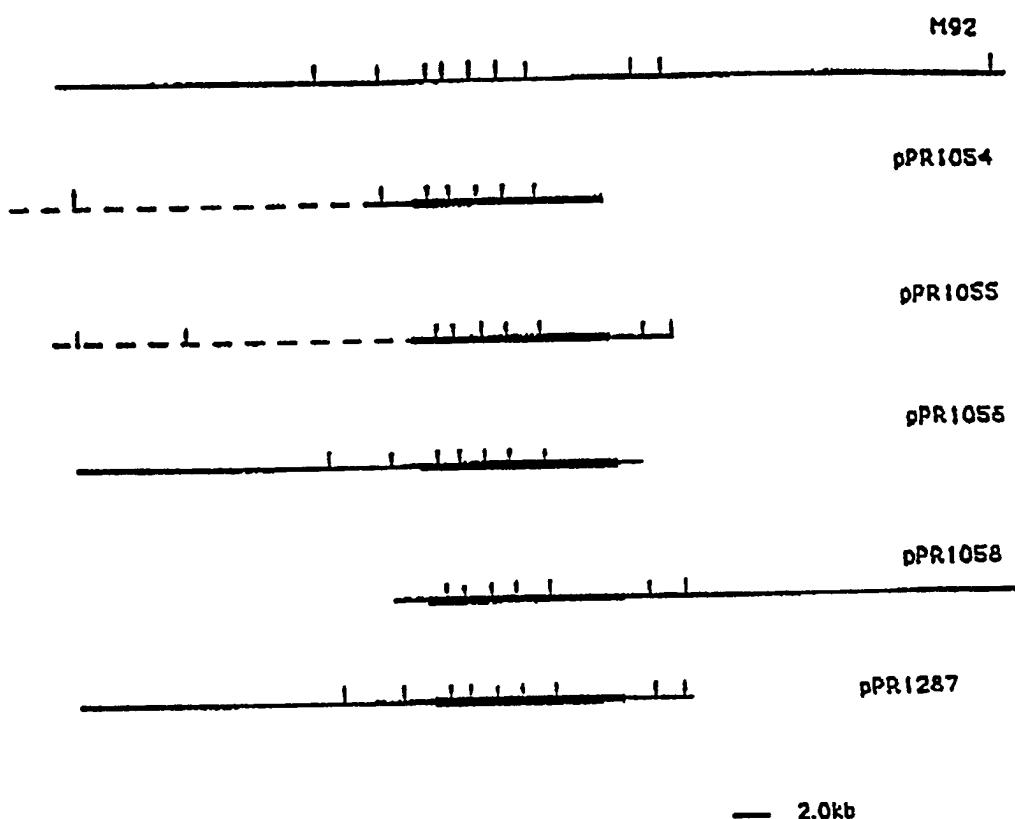


Figure 1

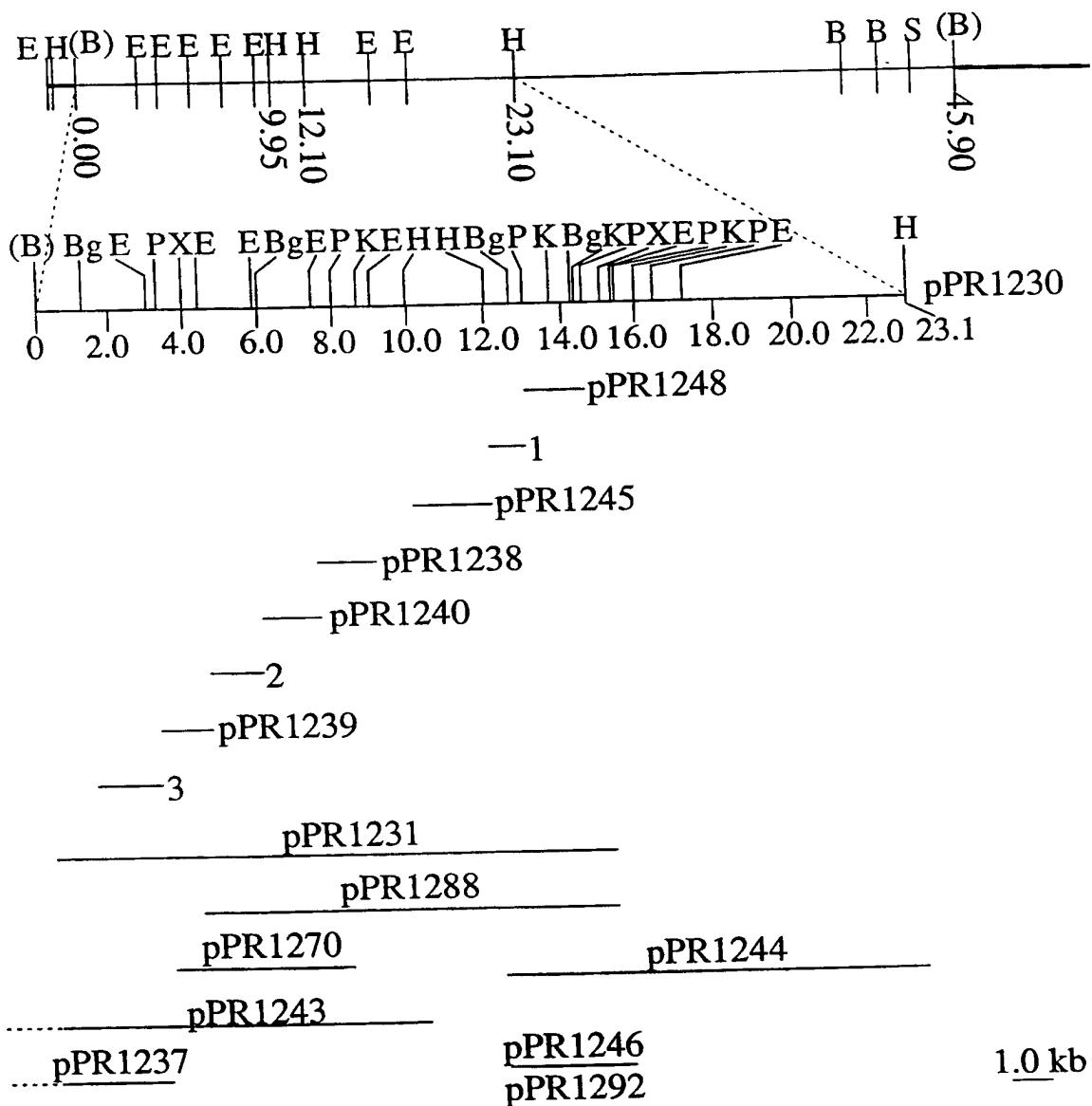


Figure 2

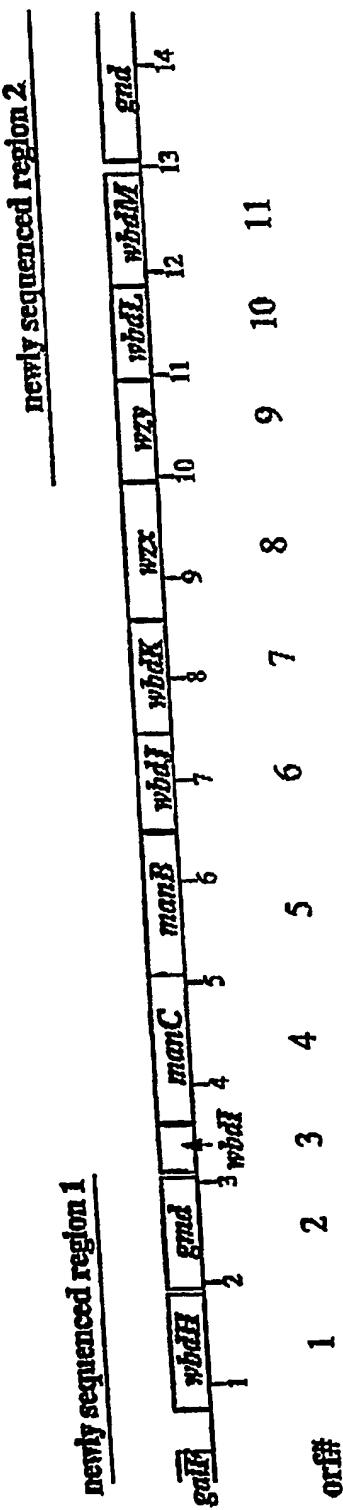


Figure 3

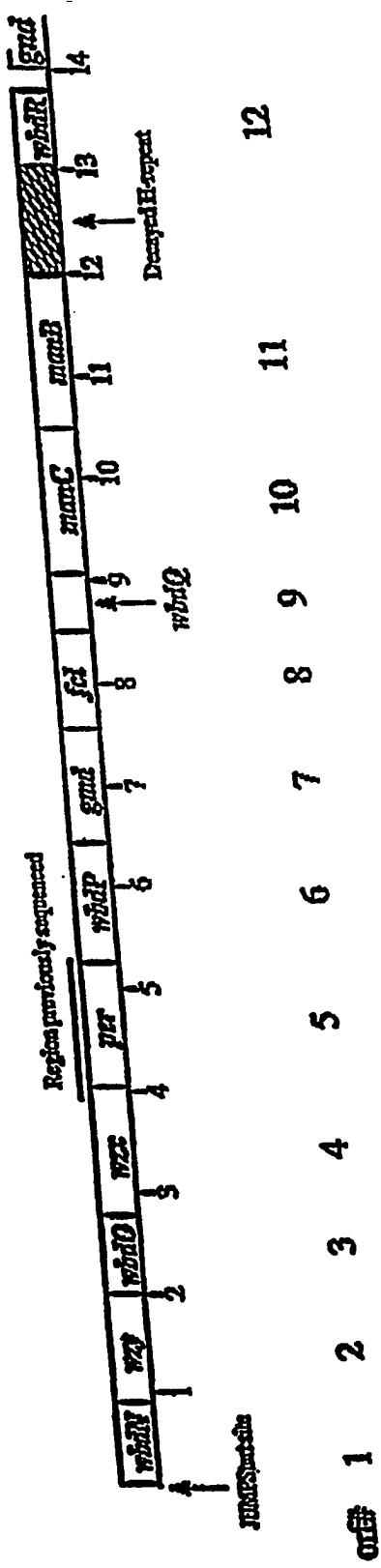


Figure 4

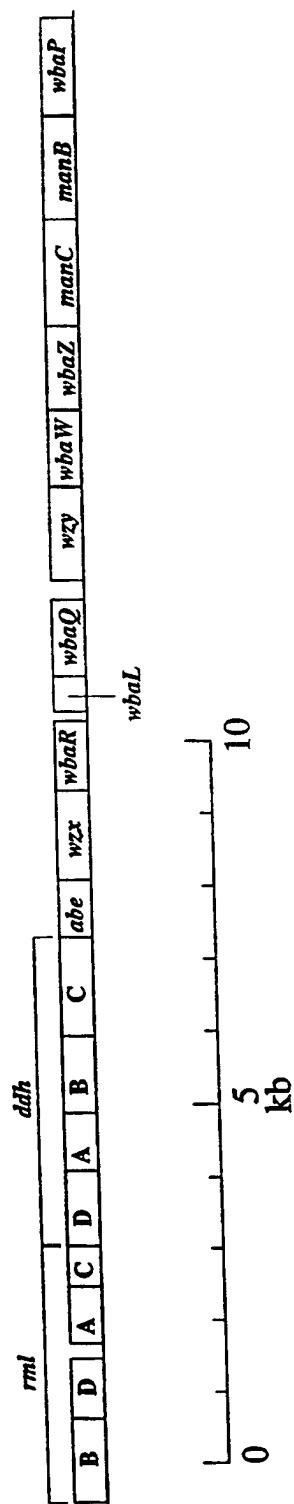


Figure 5

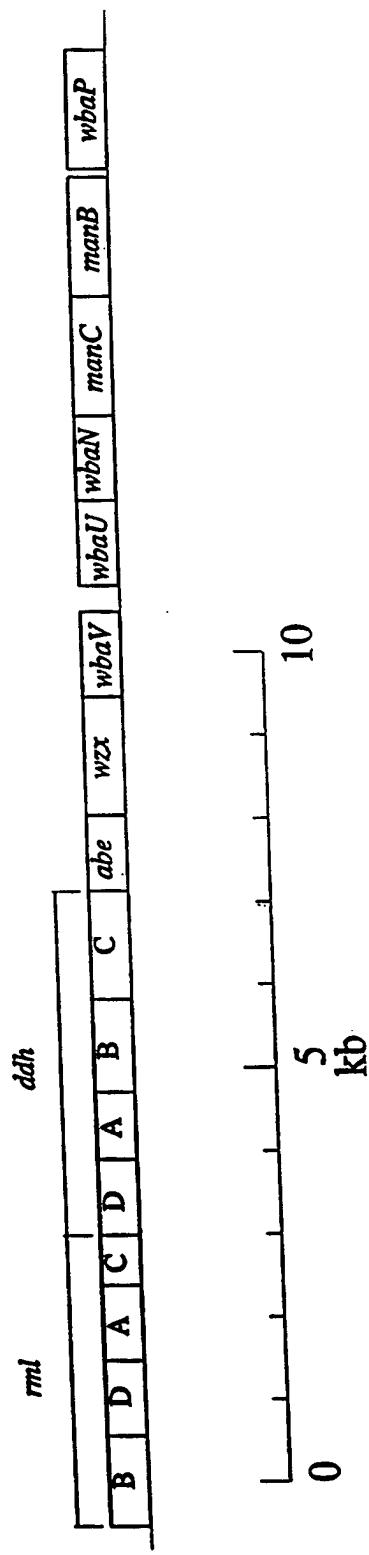


Figure 6

GATCTGATGGCCGTAGGGCGTACGTGCTTCTGCTGATATCTGGGCTGAGTTGGAAAAA	60
ACTGCTCCAGGTGCCTGGGACGTATTCAACTGACTGATGCTATTGCAGAGTTGGCTAAA	120
AAACAGTCTGTTGATGCCATGCTGATGACCGGGACAGCTACGACTGCGTAAGAAGATG	180
GGCTATATGCAGGCATTGTTAAGTATGGCTGCGAACCTTAAAGAAGGGCGAAGTTC	240
CGTAAGAGCATCAAGAAGCTACTGAGTGAGTAGAGATTACACGTCTTGTGACGATAAG	300
CCAGAAAAAATAGCGGCAGTTAACATCCAGGCTTATGCTTAAGCAATGGAATGTTAC	360
TGCCGTTTTATGAAAAATGACCAATAATAACAAGTTAACCTACCAAGTTAATCTGCT	420
TTTTGTTGGATTTCTGTTCTGGTCGCATTGGTAAGACAATTAGCGTGAGTTTA	480
GAGAGTTTGCAGGGATCTCGCGAACTGCTCACATCTTGCATTAGTTAGTGCAGTGG	540
TAGCTGTTAACGCCAGGGCGGTAGCTGCCTAACATTAAACGTATAACATTATTCT	600
TGCCGTTATAGCAAATAAGTCATCGGATTAACCTCTTTCCATTAGGTAAAAGAGT	660
GTGGTAGTCGCTCAGGGAAATTGGTTTGGTAGTAGTACTTTCAAATTATCCATTTC	720

Start of orf1

M L L C C I H I N V Y Y L L	
CGATTTAGATGGCAGTTG <u>ATGTT</u> ACTATGCTGCATAACATATCAATGTATATTATTACTT	780
L E C D M K K I V I I G N V A S M M L R	
TTAGAATGTGATATGAAAAAAATAGTGTACAGGCAATGTAGCGTCAATGATGTTAAGG	840
F R K E L I M N L V R Q G D N V Y C L A	
TTCAGGAAAGAATTAAATCATGAATTAGTGTAGGCAAGGTGATAATGTATATTGTCTAGCA	900
N D F S T E D L K V L S S W G V K G V K	
AATGATTTCCACTGAAGATCTTAAAGTACTTCGTCAAGGGCTTAAGGGGTTAAA	960
F S L N S K G I N P F K D I I A V Y E L	
TTCTCTCTTAACTCAAAGGGTATTAATCCTTTAACGGATATAATTGCTGTTATGAACTA	1020
K K I L K D I S P D I V F S Y F V K P V	
AAAAAAATTCTTAAGGATATTCCCCAGATATTGTATTTCATATTGTAAAGCCAGTA	1080
I F G T I A S K L S K V P R I V G M I E	
ATATTGGAACATTGCTCAAAGTTGTCAAAAGTGCACAGGATTGTGGAATGATTGAA	1140
G L G N A F T Y Y K G K Q T T K T K M I	
GGTCTAGGTAATGCCTTCACTTATTATAAGGGAAAGCAGACCAACAAACTAAATGATA	1200
K W I Q I L L Y K L A L P M L D D L I L	
AAGTGGATAACAAATTCTTTATATAAGTTAGCATTACCGATGCTGATGATTGATTCTA	1260
L N H D D K K D L I D Q Y N I K A K V T	
TTAAATCATGATGATAAAAAGATTTAACGATCGATCAGTATAATTAAAGCTAACGTAACA	1320
V L G G I G L D L N E F S Y K E P P K E	
GTGTTAGGTGGATTGGATCTTAATGAGTTTCATATAAGAGGCCACCGAAAGAG	1380
K I T F I F I A R L L R E K G I F E F I	
AAAATTACCTTATTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTGAGTTATT	1440
E A A K F V K T T Y P S S E F V I L G G	
GAAGCCGAAAGTTGTTAACGACAACCTATCCAAGTTCTGAATTGTAATTAGGAGGT	1500

Figure 7/1

F E S N N P F S L Q K N E I E S L R K E	1560
TTTGAGAGTAATAATCCTTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAA	
H D L I Y P G H V E N V Q D W L E K S S	1620
CATGATCTTATTATCCTGGTCATGTGGAAAATGTTCAAGATTGGTAGAGAAAAGTTCT	
V F V L P T S Y R E G V P R V I Q E A M	1680
GTGTTTGTGTTACCTACATCATCGAGAAGGCGTACCAAGGGTGTACCTGAGAAGCTATG	
A I G R P V I T T N V P G C R D I I N D	1740
GCTATTGGTAGACCTGTAATAACAACTAATGTACCTGGGTGTAGGGATATAATAATGAT	
G V N G F L I P P F E I N L L A E K M K	1800
GGGGTCAATGGCTTTGATACTCCATTGAAATTAAATTACTGGCAGAAAAATGAAA	
Y F I E N K D K V L E M G L A G R K F A	1860
TATTTTATTGAGAATAAAGATAAAGTACTCGAAATGGGCTTGCTGAGAAGGTTGCA	
E K N F D A F E K N N R L A S I I K S N	1920
GAAAAAAACTTGATGCTTTGAAAAAATAATAGACTAGCATCAATAATAAAATCAAAT	

End of orf1

N D F *

AATGATTTTGACTTGAGCAGAAATTATTTATTTCAATCTGAAAATAAAGGCTGTTA	1980
--	------

Start of orf2

M N K V A L I T G I T G Q D G S Y L A	2040
TTATGAATAAAAGTGGCATTAAATTACTGGTATCACTGGCAAGATGGCTCCTATTGGCAG	
E L L L E K G Y E V H G I K R R A S S F	2100
AATTATTGTTAGAAAAGGTTATGAAGTTCATGGTATTAAACGCCGTGCATCTTCAATTAA	
N T E R V D H I Y Q D S H L A N P K L F	2160
ATACTGAGCGAGTGGATCACATCTACAGGATTCACATTAGCTAACCTAAACTTTTC	
L H Y G D L T D T S N L T R I L K E V Q	2220
TACACTATGGCGATTTGACAGATACTTCAATCTGACCCGTATTTAAAAGAAGTTCAAC	
P D E V Y N L G A M S H V A V S F E S P	2280
CAGATGAAGTTACAATTGGGGCGATGAGCCATGTAGCGGTATTTGAGTCACCAAG	
E Y T A D V D A I G T L R L L E A I R I	2340
AATACACTGCTGATGTTGATGCGATAGGAACATTGCGTCTTCTTGAAGCTATCAGGATAT	
L G L E K K T K F Y Q A S T S E L Y G L	2400
TGGGGCTGGAAAAAAAGACAAAATTTCAGGCTTCAACTTCAGAGCTTATGGTTGG	
V Q E I P Q K E T T P F Y P R S P Y A V	2460
TTCAAGAAATTCCACAAAAGAGACTACGCCATTATCCACGTTGCCATTGCTGTTG	
A K L Y A Y W I T V N Y R E S Y G M F A	2520
CAAAATTATATGCCTATTGGATCACTGTTAATTATCGTGAGTCTTATGGTATGTTGCCT	
C N G I L F N H E S P R R G E T F V T R	2580
GCAATGGTATTCTCTTAAACCACCGAATCACCTCGCCGTGGCAGACCTTGTACTCGTA	
K I T R G I A N I A Q G L D K C L Y L G	2640
AAATAACACCGCGGGATAGCAAATTGCTCAAGGTCTGATAATGCTTATACTGGAA	
N M D S L R D W G H A K D Y V K M Q W M	2700
ATATGGATTCTCTGCGTGATTGGGACATGCTAAGGATTATGTCAAAATGCAATGGATGA	

Figure 7/2

M L Q Q E T P E D F V I A T G I Q Y S V
 TGCTGCAGCAAGAACTCCAGAAGATTTGTAATTGCTACAGGAATTCAATATTCTGTCC 2760

R E F V T M A A E Q V G I E L A F E G E
 GTGAGTTGTCACAATGGCGCAGAGCAAGTAGGCATAGAGTTAGCATTGAAAGGTGAGG 2820

G V N E K G V V V S V N G T D A K A V N.
 GAGTAAATGAAAAAGGTGTTGTTCTCGGTCAATGGCACTGATGCTAAAGCTGTAAACC 2880

P G D V I I S V D P R Y F R P A E V E T
 CGGGCGATGTAATTATCTGTAGATCCAAGGTATTTAGGCCTGCAGAAGTTGAAACCT 2940

L L G D P T N A H K K L G W S P E I T L
 TGCTTGGCGATCCTACTAATGCGCATAAAAAAATTAGGATGGAGGCCCTGAAATTACATTGC 3000

R E M V K E M V S S D L A I A K K N V L
 GTGAAATGGTAAAAGAAATGGTTCCAGCGATTTAGCAATAGCGAAAAAGAACCTTTCG 3060

End of orf2

L K A N N I A T N I P Q E *
 TGAAAGCTAATAACATTGCCACTAATATTCCGCAAGAATAAAAAGATAATACATTAAAT 3120

Start of orf3

M F
AATTAAAAATGGTGCTAGATTATTAGTACCAATTATTTTTTTGGGTGACTAATGTTTA 3180

I T S D K F R E I I K L V P L V S I D L
 TTACATCAGATAAAATTAGAGAAATTATCAAGTTACTTCCATTAGTATCAATTGATCTGC 3240

L I E N E N G E Y L F G L R N N R P A K
 TAATTGAAAACGAGAATGGTCAATATTATTTGGTCTTAGGAATAATGACCGGGCCAAA 3300

N Y F F V P G G R I R K N E S I K N A F
 ATTATTTTTTGTTCCAGGTGGTAGGATTCGAAAAATGAATCTATTAAAATGCTTTTA 3360

K R I S S M E L G K E Y G I S G S V F N
 AAAAGAATATCATCTATGAAATTAGGTAAGAGATATGGTATTTCAAGGAATGTTTTAATG 3420

G V W E H F Y D D G F F S E G E A T H Y
 GTGTATGGGACATTTCTATGATGATGGTTTTCTGAAGGGAGGGAAACACATTATA 3480

I V L C Y T L K V L K S E L N L P D D Q
 TAGTGTCTTGTACACACTGAAAGTTCTAAACTGAATTGAACTCTCCAGATGATCAAC 3540

H R E Y L W L T K H Q I N A K Q D V H N
 ATCGTGAATACCTTGGCTAACTAAACACCAAAATAATGCTAAACAAAGATGTTCAATACT 3600

End of orf3

Start of orf4

M
Y S K N Y F L *
ATTCAAAAAATTATTTTTGTAATTTTTATTAAATATGGAGAGAAATTGTT 3660

S Q C L Y P V I I A G G T G S R L W P L
 etcAAATGTCTTACCTGTAATTATTGGGGAGGGAAACGGGAGGGAACTGCTATGGGGTTGT 3720

S R V L Y P K Q F L N L V G D S T M L Q
 etcGACTATTATACCCCTAAACAAATTATTAGTTGGGGATTCTACAAATGTTCAAA 3780

T T I T R L D G I E C E N P I V I C N E
 CAACAAATTACGGCTTGGATGGCATCGAATGCGAAAATCCAATTGTTATCTGCAATGAG 3840

D H R F I V A E Q L R Q I G K L T K N I
 ATCAACGGATTATTGTAACGAGAGCAATTACGGACAGATTGTAACGTAACCAAGAAATTAA 3900

I L E P K G R N T A P A I A L A A F I A
 TACTTGAGCGAAAGGCGTAAATGCAACCTGCAACCTGTTAGCTGTTTATGCTG 3960

Figure 7/3

10/58

Q K N N P N D D P L L L V L A A D H S I	4020
AGAAGAATAATCCTAATGAAAGACCCCTTATTATAGTACTTGGGGAGACCACTCTATAA	
N N E K A F R E S I I K A M P Y A T S G	4080
ATAATGAAAAAGCATTTCGAGACTCAATAAAGCTATGCCGTATGCAACTTCTGGGA	
K L V T F G I I P D T A N T G Y G Y I K	4140
AGTTAGTAACATTGGAATTATTCGGACACGGCAAATACTGGTTATGGATATATTAAGA	
R S S S A D P N K E F P A Y N V A E F V	4200
GAAGTTCTTCAGETGATECTAATAAAGAATTCGGACATATAATGTTGGGAGTTGTAG	
E K P D V K T A Q E Y I S S G N Y Y W N	4260
AAAAACAGATGTTAACACGACAGGAATATATTCGAGTGGAAATTATTACTGGAAATA	
S G M F L F R A S K Y L D E L R K F R P	4320
CGGAAATGTTTTATTCGGCCAGTAATACTTGATGAACTACGGAAATTAGACCAAG	
D I Y H S C E C A T A T A N I D M D F V	4380
ATATTATCATAGCTGTGAAATGCAACCGTACAGCAAATATAGATATGGACTTTGTC	
R I N E A E F I N C P E E S I D Y A V M	4440
GAATTAAACGAGGCTGAGTTATTATTGTECTGAAGACTCTATCGATTATGCTGTGATGG	
E K T K D A V V L P I D I G W N D V G S	4500
AAAAAACAAAACACCGCTGAGTTCTTCCGATAGATATTGGCAACGTTGGGTCTT	
W S S L W D I S Q K D C H G N V C H G D	4560
GGTCATCACTTGGGATATAACCCAAAAGGATTGCCATGCTAATGCTGCAATGGGATG	
V L N H D G E N S F I Y S E S S L V A T	4620
TCCTCAATCATGATGGAGAAAATAGTTTATTACTCTGAGTCAGTCGTTGGCACCG	
V G V S N L V I V Q T K D A V L V A D R	4680
TCGGASTAAGTAATTAGTAATTGTECAACCAAGGATGCTGTTGGTTGGGACCGTG	
D K V Q N V K N I V D D L K K R K R A E	4740
ATAAAAGTCCAAAATGTTAACATAGTTGACGATCTAAAAGAGAAAACGTGCTGAAT	
Y Y M H R A V F R P W G K F D A I D Q G	4800
ACTACATGCACTGTCAGTTTCCCGTTGGCTAAATGCTGCAATAGACCAAGGCG	
D R Y R V K K I I V K P G E G L D L R M	4860
ATAGATATAGAGTAACCAAGGAGAAGGGTAGATTTAACGATG	
H H H R A E H W I V V S G T A K V S L G	4920
ATCATGATAGGGCAGAGCATGGATTGTTGCTATCCCTACTGCTAAAGTTCACTAGGTA	
S E V K L L V S N E S I Y I P Q G A K Y	4980
CTGAAGTTAACTATTAGTTCTAATGAGTCATATATATCCCTCAGGGAGCAAAATATA	
S L E N P G V I P L H L I E V S S G D Y	5040
CTCTTGAGAATCCACCCGTAATACTTTGCACTAATGAACTAAGTCTCTGATTAC	
L E S D D I V R F T D R Y N S K Q F L K	5100
TTGAATCAGATGATATAAGTGCCTTTACTGACAGATATAACAGTAAACAAATTCTAAAGC	

End of orf4 Start of orf5
 M N K I T C F K A Y D I R G R L

R D *
GAGATTGATAATAATGAAATAAAACTTGCTCAAGGATATGATATAACGTGGGGCT 5160

G A E L N D E I A Y R I G R A Y G E F F 5220
 TGGTGCCTGAATTGAAATGATGAAATAGCATATAGAATTGCTTGGCTTATGGTGAGTTTT
 K P Q T V V V G G D A R L T S E S L K K 5280
 TAAACCTCAAACGTACTTGTGGGAGGAGATGCTGCTTAACAACTGAGAGCTTTAAAGAA
 S L S N G L C D A G V N V L D L G M C G 5340
 ATCACTCTCAAATGGGCTATGTGATGCAGGCTAAATGCTTAGATCTTGGAAATGTGTGG
 T E E I Y F S T W Y L G I D G G I E V T 5400
 TACTGAAGAGATATATTTTCCACTTGGTATTAGGAATTGATGGTGGAAATCGAGGTAAC
 A S H N P I D Y N G M K L V T K G A R P 5460
 TGEAAGCCATAATCCAATTGATTATAATGAAATTGAAATTAGTAACCAAAGGTGCTGAC
 I S S D T G L K D I Q Q L V E S N N F 5520
 AATCAGGACTGACACAGCTCTCAAAGATATACACAAATTAGTAGAGAGTAATAATTGTGA
 E L N L E K K G N I T K Y S T R D A Y I 5580
 AGAGETCAACCTAGAAAAAAACCAATTACCAAATTCTACACCGAGATGCTTACAT
 N H L M G Y A N L Q K I K K I K I V V N 5640
 AAATCAATTGATGGGCTATGCTAATCTGCAAAATAAAAAAAATCAAATTAGTTGTGAA
 S G N G A A G P V I D A I E E C F L R N 5700
 TTCTGGGAATTGTCAGCTTGTGCTTATTGATGCTATTGAGGAATTGTTTTACGGAA
 N I P I Q F V K I N N T P D G N F P H G 5760
 CAATATTCCGATTCACTTGTAAAAATAATACACCCGATGCTATTTCGACATGG
 I P N P L L P E C R E D T S S A V I R H 5820
 TATCCCTAAATCCATTACTACCTGAGTGGAGAGAAGATACCGACTGGGTTATAAGACA
 S A D F G I A F D G D F D R C F F F D E 5880
 TAGTGCTGATTTGGTATTGCAATTGATGGTGTGATAGCTGTTTTCTTGATGA
 N G Q F I E G Y Y I V G L L A E V F L G 5940
 AAATGGACAATTATTGAAAGGATACTACATTGGTTATTAGGGAAAGTTTTTACGG
 K Y P N A K I I I H D P R L I W N T I D I 6000
 GAATATCCAACCCAAAAATCATGATGCTEGEETATATGAAATTACTATTGATAT
 V E S H G G I P I M T K T G H A Y I K Q 6060
 CCTAGAAACTCATGGTGGTATAACCTATAATGACTAAACCGGTCTGCTTACATTAGCA
 R M R E E D A V Y G G E M S A H H Y F K 6120
 AGAAATGGTGAAGAGGATGGCTATATGGGGGAAATGAGTCGGCATEATTATTAA
 D F A Y C D S G M I P W I L I C E L L S 6180
 AGATTTGCTACTGGATAGTGGATTGATTECTTGGATTAAATTGTCAGTTTGAG
 L T N K K L G E L V C G C I N D W P A S 6240
 TETGACAAATTAGCTGAACTGGTTGTGGTTGTATAACGACTGGGGGAAAG
 G E I N C T L D N P Q N E I D K L F N R 6300
 TCGAGAAATAACTGTACACTAGACAATCCGAAAAATGAAATTAGATAATTATTAATCG
 Y K D S A L A V D Y T D G L T M E F S D 6360
 TTACAAAGATACTGECCTAGCTGTTGATTACACTGATGGATTAACTATGGAGTTCTCTGA
 W R F N V R C S N T E P V V R L N V E S 6420
 TTGGCCTTTAATGTTAGATGCTAAATACAGAACCTGTACTACGATTGAAATGTAGAAATE
 R N N A I L M Q E K T E E I L N F I S K 6480
 TAGGAATAATGCTATTCTTATGGAGAAAAACAGAAGAAATTCTGAAATTATATCAA

Figure 7/5

12/58

Start of orf6

M K V L L T G
6540

End of orf5

ATAAATTTCGACCTGAGTTCAATACTGGAAACAAGAAATATATGAAAGTACTTCTGACTGG
6600

S T G M V G K N I L E H D S A S K Y N I
CTCAACTGGCATGGTTGGTAAGAAATATATTAGACCATGATAGTGCAGAAGTAAATATAATAT
6660

L T P T S S D L N L L D K N E I E K F M
ACTTACTCCAAACCAACETGTGATTGAAATTATTAGATAAAAATGAAATAGAAAAATTCAAT
6720

L I N M P D C I I H A A G L V G G I H A
GCTTATCAACATGCCAGACTGTATTATACATGCAGGGGATTAGTTGGAGGGATTCATGC
6780

N I S R P F D F L E K N L Q M G L N L V
AAATATAAGCAGGGCGTTGATTTCTGGAAAAAAATTGAGATGCGTTAAATTAGT
6840

S V A K K L G I K K V L N L G S S C M Y
TTCGGTGGCAAAACTAGGTATAAGAAAGTGCCTAACGGGTAGTTCATGCATGTA
6900

P K N F E E A I P E K A L L T G E L E E
CCCCAAAACCTTGAGAGGCTATTCTGAGAAAGCTCTGTTAACGGTGAAGCTAGAAGA
6960

T N E G Y A I A K I A V A K A C E Y I S
AACTAATGAGGGATATGCTATTGCGAAAATTGCTGTAGCAAAACCATGCGAAATATATC
7020

R E N S N Y F Y K T I I P C N L Y G K Y
AAGAGAAAACCTCAATTATTTTATAAAACATTATCCATGTAATTATATGGAAATA
7080

D K F D D N S S H M I P A V I K K I H H
TGATAAAATTGATGATAACTCTGACATATGATCCGGGAGTTATAAAACCTCA
7140

A K I N N V P E I E I W G D G N S R R E
TCCGAAATTAAATGCTTACAGAGATCGAAATTGGGGGATGCTTAATCGGGCGTGA
7200

F M Y A E D L A D L I F Y V I P K I E F
GTTTATGTTATGCAAGAAGATTGCTGATCTTATTTTATGTTATTCTAAATAGAATT
7260

M P N M V N A G L G Y D Y S I N D Y Y K
CATGCCATAATGGTAAATGCTGTTACGTTACGATTTCATAATGACTATTATAA
7320

I I A E E I G Y T G S F S H D L T K P T
GATAATTGCGAGAAGAAATTGGTTACTGGAGTTTCTCATGATTAACAAAACCAAG
7380

G M K R K L V D I S L L N K I G W S S H
AGGAATGAAACCGAACCTAGTAGATATTGCTTAATAAAATTGCTGGTCAGTC
7440

F E L R D G I R K T Y N Y Y L E N Q N K
CTTGTAACTCAGAGATGGCATCAGAAAGACCTATAATTACTGGAGAAATCAAATAA

Start of orf7, End of orf6

M I T Y P L A S N T W D E Y E Y A A I Q
7500

ATGATTACATAACCACTTGCTAGTAATACTGGCATGAAATATGACTATGCAGCAATACAG

S V I D S K M F T M G K K V E L Y E K N
TCAGTAATTGACTCAAAATGTTACCATGGCTAAAAGGTTGACTTATATGAGAAAT
7560

F A D L F G S K Y A V M V S S G S T A N
TTTGTCAATTGTTGCTAGCAAAATATGCGCTAATGCGTTAGCTCGGTTCTACAGCTAA
7620

Figure 7/6

L L M I A A L F F T N K P K L K R G D E
 CTGTTAATGATTGCTTCCCTTTCTTCACTAATAACCAAAACTTAAAGAGGTGATGAA 7680
 I I V P A V S W S T T Y Y P L Q Q Y G L
 ATAATAGTACCTGCACTGTCATGGTCTACGACATATTAACTCTGCAACAGTATGGCTTA 7740
 K V K F V D I N K E T L N I D I D S L K
 AAGGTGAAGTTTGTGATATCAATAAGAAACTTAAATATTGATATCGATACTTGTAAA 7800
 N A I S D K T K A I L T V N L L G N P N
 ATGCTATTCAGATAAAAGAAATATTGACAGTAATTTATTAGGTAATCTTAAT 7860
 D F A K I N E I I N N R D I I L L E D N
 GATTTGCAAAATAATGAGATAATAATAAGGGATATTATCTTACTAGAAGATAAC 7920
 C E S M G A V F Q N K Q A G T F G V M G
 TGTGAGTCGATGGGGGGGCTTTCATAAAAGCAGGGCACATTGGACTTATGGGT 7980
 T F S S F Y S H H I A T M E G G C V V T
 ACCTTTAGTTCTTACTCTCATATAGCTACAATGGAAAGGGGCTGGCTAGTTACT 8040
 D D E E L Y H V L L C L R A H G W T R N
 GATGATGAAGAGCTGTATCATGTATGTTGCTTCTGAGCTCATGGTGGACAACAAAT 8100
 L P K E N M V T G T K S D D I F E E S F
 TTACCAAAAGAGAAATATGTTACAGGCACTAAGACTGATGATATTTCGAAGAGCTGGTT 8160
 K F V L P G Y N V R P L E M S G A I G I
 AAGTTTGTCTTACCAAGGATACAATGTTGGCCACTTGAATGACTGCTGCTATTGGATA 8220
 E Q L K K L P G F I S T R R S N A Q Y F
 GACCAACTTAAAGACTTACCAAGGTTATATECACCAGACCTTCCATGGCACAAATTTT 8280
 V D K F K D H P F L D I Q K E V G E S S
 GTAGATAAAATTAAAGATCATECCATTCTTGATATACAAGAAAGTTGGTGAAGTAC 8340
 W F G F S F V I K E G A A I E R K S L V
 TGTTTGGTTCTTCTGTTATAAAGGAGGGAGCTGTTATTGAGAGGAAGAGTTAGTA 8400
 N N L I S A G I E C R P I V T G N F L K
 ATAATCTGATCTGAGGCAATTGAATGGCCAAATTGTTACTGGGAATTTCCTCAAA 8460
 N E R V L S Y F D Y S V H D T V A N A E
 ATGAACTGTTTGACTTATTTGATTACTCTGATACGGTAGCAAATGGCAAA 8520
 Y I D K N G F F V G N H Q I P L F N E I
 TATATAGATAAGAATGGTTTTGCTGGAAACCAACGATAACCTTGTAAATGAAATA 8580

End of orf7
 D Y L R K V L K *
 GATTATCTACGAAAAGTATTAAATACTAACGAGGCACTCTATTGAAATAGCTGGCT 8640

Start of orf8
 M V L T V K K I L A F G Y S K V L P
 TTAAGATGGTATAACACTGAAAGAAATTCTACGGTTGGCTATTCTAAAGTACTACCA 8700
 P V I E Q F V N P I C I F I I T P L I L
 EGTTATTGAAACAGTTGTCAATCCATTGCTATTGATCTTCAACACCACTAACTCA 8760
 N H L G K Q S Y G N W I L L I T I V S F
 ACCACCTGGCTAACCAAGCTATGGTAATTGGATTTTATTAACTATTGTATCTTT 8820

Figure 7/7

S Q L I C G G C S A W I A K I I A E Q R	8880
CTCAAGTTAATATGTGGAGGATGTTGGCATGGATTGCAAAATCATGGAGAACAGAGAA	
I L S D L S K K N A L R Q I S Y N F S I	8940
TTCTTACTGATTATCAAAAAAAATGCTTACGTCAAATTCTATAATTTCATTCATTG	
V I I A F A V L I S F L I L S I C F F D	9000
TTATTATGGCATTTGGGCTATTGATTCTTTCTTATATTAACTATTGTTCTCGATG	
V A R N N S S F L F A I I I C G F F Q E	9060
TTGGGAGGAATAATTCTCATTCATTCTTATGGCATTATTATTTGTGCTTTTTTCAGGAAG	
V D N L F S G A L K G F E K F N V S C F	9120
TTGATAATTATTTAGTGGTGGCTTCTATAGTAATATATGGCATTTACGGAAATG	
F E V I T R V L W A S I V I Y G I Y G N	9180
TTGAACTTAATTACAAGAGTGTCTGGCTTCTATAGTAATATATGGCATTTACGGAAATG	
A L L Y F T C L A F T I K G M L K Y I L	9240
CACTCTTATATTTCACATGTTAGCCTTACCAATTAAAGCTATGTAATATTATTCATG	
V C L N I T G C F I N P N F N R V G I V	9300
TATGTCATGAATATTACCGGTTGTTCATCAATCTAATTAAAGACTTGGATTGTTA	
N L L N E S K W M F L Q L T G G V S L S	9360
ATPTGTTAAATGAGTCAAATGGATGTTCTCAATTAACTGGCTTCTCACTTACTTACTT	
L F D R L V I P L I L S V S K L A S Y V	9420
TETTTGATAGCTTCTAATACCATGATTTTATCTGTCACTAACTGGCTTATGTCTT	
P C L Q L A Q L M F T L S A S A N Q I L	9480
CTTGGCTTCAACTAGCTCAATTGATGTTCACTCTTCTGGCTTCTGAAATCAATATTAG	
L P M F A R M K A S N T F P S N C F F K	9540
TACCAATGTTGCTAGAATGAAACCATCTAACACATTCTCTCTAAATTGTTTTAAA	
I L L V S L I S V L P C L A L F F F G R	9600
TTCTGTTGATCACTAATTCTCTTCTGGCTTCTGGTTATCTTTCTGCTCTG	
D I L S I W I N P T F A T E N Y K L M Q	9660
ATATATTATCAATATGGATAACCTACATTGCAACTGAAATTATAAAATTAAATGCAA	
I L A I S Y I L L S M M T S F H F L L L	9720
TTTTAGCTATAAGTACATTGCAATTGATGACATCTTCTGTTATTAG	
G I G K S K L V A N L N L V A G L A L A	9780
GAATTGGTAATCTAACCTGGCTTCAATTAACTGGCTTCTGGCTTCTGCTG	
A S T L I A A H Y G L Y A I S M V K I I	9840
ETTCACCGTTAACCGAGCTCATPATGGCTTATGCAATTATCTATGCTAAATAATT	
Y P A F Q F Y Y L Y V A F V Y F N R A K	9900
ATCCCCCTTCAATTATTACCTTATGAGCTTGTCTATTAAATGAGGAA	

Start of orf9, End of orf8

M S I D L L F S I T E I A I V F S C T I	
N V Y *	
<u>ATGTCTATTGATTCTCTTCAATTACTGAAATCGCAATTGTTCTGCACTATT</u>	9960
Y I F T O C L L M R R I Y L D K S I L I	10020
TACATATTACTCAATTGTTGTTAATGCGGAGGATCTATTAGATAAAAGTATTAAATT	
L L C L L F F L V I I Q L P E L N V N G	10080
CTTTATGCTTGCCTTTAGTAATCAACTCCTGAGCTTAATGAAACGGT	

Figure 7/8

L V D S L K L S L P L L M V F I A F Q K		
TTGGTCGATTCTTAAAGTTATCACTGCCTTATTGATGGCTTATCGCTTCAAAAAA	10140	
P K L C L W V I I A L L F L N S A F N F		
CCGAAATTATGCTTGTGGTTATTGCATTGTTTTGAACCTCTGCATTAAATT	10200	
L Y L K T F D K F S S F P F T F F I L L		
TTATATTAAAGACATTCGATAAGTTAGCTATTCCCTTTACTTTTTATATTGCTG	10260	
F Y L F R L G I G N L P V Y K N K K F Y		
TTTACTTGTAGATTGGGAATTGGAATTACGGTTATAAAAATAAAATTAC	10320	
A L I F L F I L I D I M Q S L L I N Y R		
GCCTGATTTCTTTATTAATAGACATAATGCAGTCATTGTTAATAATTAGG	10380	
G Q I L Y S V I C I L I L V F K V N L R		
GGCAGATTTATTCCTGTAATTGCATCCTGATACTTGTGTTAAAGTTAATTAAAGA	10440	
K K I P Y F F L M L P V L Y V I I M A Y		
AAAAGATTCCATACTTTTTAATGCTGCCAGTTATGTAATTATTATGGCTTAT	10500	
I G F N Y F N K G V T F F E P T A S N I		
ATTGGTTTAATTATTCAATAAGCGTAACCTTTGAACTACAGCAAGTAATTATT	10560	
E R T G M I Y Y L V S Q L G D Y I F H G		
GAACGTACGGGGATGATATTATTGGTTACAGCTGGTATTATATTCCATGGT	10620	
M G T L N F L N N G G Q Y K T L Y G L P		
ATGGGGACATTAATTCCTAAACGGCGACAATAAGACGTTATATGGACTTCA	10680	
S L I P N D P H D F L L R F F I S I G V		
TCATTAATTCTAATGACCTCATGATTTTATTACGGTTCTTATAAGTATTGGTGTG	10740	
I G A L V Y H S I F F V F F R R I S F L		
ATAGGAGCATTGGTTATCATTCTATATTGGTTAGGAGAATATCTTCTTA	10800	
L Y E R N A P F I V V S C L L L Q V V		
TTATATGAGAGAAATGCTCCTTCATTGTTAAGTTGTTACTGTTACAAGTGTG	10860	
L I Y T L N P F D A F N R L I C G L T V		
TTAATTATACATTAACCCCTTTGATGCTTTAATCGATTGATTGCGGGCTTACAGTT	10920	

Start of orf10	End of orf9	
G V V Y G F A K I R *		
M D L Q K L D K Y T C N G N L D A		
GGAGTTGTTATGGATTGCAAAATTAGATAAGTATACTGTAATGAAATTAGACGC	10980	
P L V S I I I A T Y N S E L D I A K C L		
TCCACTTGTTCATAATCATTGCAACTTATAATTCTGAACTTGATATAGCTAAGTGT	11040	
Q S V T N Q S Y K N I E I I I M D G G S		
GCAATCGGTAACATCAATCTTATAAGAATTGAAATCATATAATGGATGGAGGATC	11100	
S D K T L D I A K S F K D D D R I K I V S		
TTCTGATAAAACGCTTGTATTGCAAATCGTTAAAGACGACCGAATAAAATAGTT	11160	
E K D R G I Y D A W N K A V D L S I G D		
AGAGAAAGATCGTGGAAATTATGATGCCTGGAATAAGCAGTTGATTATCATTGGTGA	11220	
W V A F I G S D D V Y Y H T D A I A S L		
TTGGGTAGCATTTATTGGTTCAAGATGATGTTACTATCATACAGATGCAATTGCTTCATT	11280	
M K G V M V S N G A P V V Y G R T A H E		
GATGAAGGGGTTATGGTATCTAATGGCGCCCTGTGGTTATGGGAGGACAGCGCACGA	11340	

Figure 7/9

16/58

G P D R N I S G F S G S E W Y N L T G F
 AGGTCCCGATAGGAACATATCTGGATTTCAAGCAGTGAATGGTACAACCTAACAGGATT 11400

 K F N Y Y K C N L P L P I M S A I Y S R
 TAAGTTAATTATTACAAATGTAATTACCATGCCATTATGAGCGCAATATATTCTCG 11460

 D F F R N E R F D I K L K I V A D A D W
 TGATTTCTTCAGAACGAAACGTTGATATTAAATTAAAATTGTTGCTGACGCTGATTG 11520

 F L R C F I K W S K E K S P Y F I N D T
 GTTTCTGAGATGTTCATCAAATGGAGTAAAGAGAAGTCACCTTATTTATTAATGACAC 11580

 T P I V R M G Y G G V S T D I S S Q V K
 GACCCCTATTGTTAGAATGGATATGGGGTTTCGACTGATATTCTCTCAAGTTAA 11640

 T T L E S F I V R K K N N I S C L N I Q
 AACTACGCTAGAAAGTTTCATTGTACGCAAAAGAATAATATATCCTGTTAAACATACA 11700

 L I L R Y A K I L V M V A I K N I F G N
 GCTGATTCTTAGATATGCTAAAATTCTGGTAGGGATCAAAATATTGGCAA 11760

 N V Y K L M H N G Y H S L K K I K N K I
 TAATGTTATAAATTAATGCATAACGGGTATCATTCCCTAAAGAAAATCAAGAATAAAAT 11820

Start of orf11, End of orf10
 M K I V Y I I T G L T C G G A E H L M T
 *
ATGAAAGATTGTTATATAATAACCGGGCTTACTTGTGGTAGGCCAACACCTTATGACG 11880

 Q L A D Q M F I R G H D V N I I C L T G
 CAGTTAGCAGACCAATGTTATACGCGGGCATGATGTTAATATTATTGTCTAACTGGT 11940

 I S E V K P T Q N I N I H Y V N M D K N
 ATATCTGAGGTAAAGCCAACACAAAATATTAAATATTCAATTATGTTAATATGGATAAAAAT 12000

 F R S F F R A L F Q V K K I I V A L K P
 TTTAGAAGCTTTAGAGCTTATTCAGTAAAGTAAAAAAATAATTGTCGCCTAAAGCCA 12060

 D I I H S H M F H A N I F S R F I R M L
 GATATAATACATAGTCATATGTTCATGCTAATATTAGTCGTTTATTAGGATGCTG 12120

 I P A V P L I C T A H N K N E G G N A R
 ATTCCAGCGGTGCCCTGATATGACCGCACACAACAAAATGAAGGTGGCAATGCAAGG 12180

 M F C Y R L S D F L A S I T T N V S K E
 ATGTTTGTATCGACTGAGTGTAGTTAGCTTACTACAAATGTAAGTAAAGAG 12240

 A V Q E F I A R K A T P K N K I V E I P
 GCTGTTCAAGAGTTATAGCAAGAAAGGCTACACCTAAAATAAGTAGAGATTCCG 12300

 N F I N T N K F D F D I N V R K K T R D
 AATTTTATTAATACAAATAATTGATTTGATATTATGTCAGAAAGAAAACGCGAGAT 12360

 A F N L K D S T A V L L A V G R L V E A
 GCTTTAATTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTGAAGCA 12420

 K D Y P N L L N A I N H L I L S K T S N
 AAAGACTATCCGAACCTATTAAATGCAATAATCATTGATTCTTCAAAACATCAAAT 12480

 C N D F I L L I A G D G A L R N K L L D
 TGTAATGATTTATTCGCTTATTGCTGGCGATGGCGCATTAAGAAATAATTATTGGAT 12540

 L V C Q L N L V D K V F F L G Q R S D I
 TTGGTTGTCAATTGAATCTGTGGATAAAAGTTCTCTGGGGCAAAGAAGTGTATT 12600

Figure 7/10

K E L M C A A D L F V L S S E W E G F G
AAAGAATTAATGTGTGCTGCAGATCTTTGTTGAGTTCTGAGTGGAAAGGTTTGGT 12660

L V V A E A M A C E R P V V A T D S G G
CTCGTTGTTGCAAGCTATGGCGTGTGAACGTCCCGTTGCTACCGATTCTGGTGG 12720

V K E V V G P H N D V I P V S N H I L L
GTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTG 12780

A E K I A E T L K I D D N A R K I I G M
GCAGAGAAAATCGCTGAGACACTTAAATAGATGATAACGCAAGAAAATAATAGGTATG 12840

K N R E Y I V S N F S I K T I V S E W E
AAAAATAGAGAATATATGTTCCAATTTCATTAAACGATAAGTGAGTGAGTGGGAG 12900

End of orf11

R L Y F K Y S K R N N I I D *
CGCTTATATTAAATATTCCAAGCGTAATAATATAATTGAT TGAAAATATAAGTTGTA 12960

CTCTGGATGCAATAGTTCTCTATGCTGTTTTACTGGCTCCGTATTTTACTTATAG 13020

CTGGATTTGTTATATCAGTATTAAATCTGCTCAACTCATCTAGACTACATTCAAGC 13080

Start of gnd

M S K Q Q I
CGCGCATGCGTCGCGCGGTGACTACACCTGACAGGAGTATGTA ATGTCAGCAACAGAT 13140

G V V G M A V M G R N L A L N I E S R G
CGGCCTCGTCGGTATGGCAGTGATGGGGCGAACCTGGCGCTAACATCGAAAGCCGG 13200

Y T V S I F N R S R E K T E E V V A E N
TTATACCGTCTCCATCTCAACCGCTCCCGAGAAAATGAGAAGTTGTTGCCGAGAA 13260

P D K K L V P Y Y T V K E F V E S L E T
CCCGGATAAGAAAATGGTTCTTATTACACGGTGAAAGAGTTCGAGTCCTTGAAAC 13320

P R R I L L M V K A G A G T D A A I D S
CCACGTCGTATCCTGTTAATGGTAAAGCAGGGCGGAACTGATGCTGCTATCGATT 13380

L K P Y L D K G D I I D G G N T F F Q
CCTGAAGCCGTATCTGGATAAAGCGACATCATTATTGATGGTGGCAACACCTTCTTCA 13440

D T I R R N R E L S A E G F N F I G T G
GGACACTATCCGTCGTAACCGTGAACGTCCCGGAAGGCTTTAACTTCATCGGTACCG 13500

V S G G E E G A L K G P S I M P G G Q K
CGTGTCCGGCGGTGAAGAGGGCCCGTGAAAGGCCATCTATGCCAGGTGGCCAGAA 13560

E A Y E L V A P I L T K I A A V A E D G
AGAAGCGTATGAGCTGGTTGCGCCTATCCTGACCAAGATTGCTGCGGTTGCTGAAGATGG 13620

E P C I T Y I G A D G A G H Y V K M V H
CGAACCATGTATAACTACATCGGTGCTGACGGTGGGGTCACTACGTGAAGATGGTGC 13680

N G I E Y G D M Q L I A E A Y S L L K G
CAACGGTATCGAATATGGCGATATGCAGCTGATTGCTGAAGCCTATTCTGCTTAAAGG 13740

G L N L S N E E L A T T F T E W N E G E
CGGCCTTAATCTGTCTAACGAAGAGCTGGCAACCACCTTACCGAGTGGATGAAGGGCA 13800

L S S Y L I D I T K D I F T K K D E E G
GCTAAGTAGCTACCTGATTGACATCACCAAGACATCTCACCAAAAAAGATGAAGAGGG 13860

Figure 7/11

K Y L V D V I L D E A A N K G T G K W T TAAATACCTGGTTGATGTGATCCTGGACGAAGCTGCGAACAAAGGCACCGTAAATGGAC	13920
S Q S S L D L G E P L S L I T E S V F A CAGCCAGAGCTCTGGATCTGGGTGAACCGCTGTCGCTGATCACCGAATCCGTATTCGC	13980
R Y I S S L K D Q R I A A S K V L S G P TCGCTACATCTCTCTGAAAGACCAGCGCATTCGGCATCTAAAGTGTCTGGTCC	14040
Q A K L A G D K A E F V E K V R R A L Y GCAGGCTAAACTGGCTGGTGATAAAGCAGAGTTCGTTGAGAAAGTCCGTCGCGCTGTA	14100
L G K I V S Y A Q G F S Q L R A A S D E CCTGGTAAAATCGTCTTTATGCCAAGGCTCTCAACTGCGTGCCGCGTCTGACGA	14160
Y N W D L N Y G E I A K I F R A G C I I ATACAACGGGATCTGAACGCGAAATCGCGAAGATCTCCGCGCGGCTGCATCAT	14220
R A Q F L Q K I T D A Y A E N K G I A N TCGTGCGCAGTCCTGCAGAAAATTACTGACGCGTATGCTGAAACAAAGGCATTGCTAA	14280
L L L A P Y F K N I A D E Y Q Q A L R D CCTGTTGGCTCCGTACTTCAAAAATATCGCTGATGAATATCAGCAAGCGCTGCGTGA	14340
V V A Y A V Q N G I P V P T F S A A V A TGTAGTGGCTTATGCTGTCAGAACGGTATTCCGGTACCGACCTCTCTGCAGCGGTAGC	14400
Y Y D S Y R S A V L P A N L I Q A Q R D CTACTACGACAGCTACCGTTCTGCGGTACTGCCGGCTAATCTGATTCAAGGCACAGCGTGA	14460
Y F G A H T Y K R T D K E G V F H T G TTACTTCGGTGCACACGTATAAACGCACTGATAAAGAAGGTGTGTTCCACACCG	14516

GTAACCAAGGGCGGTACGTGCATAAATTTAATGCTTATCAAAACTATTAGCATTAAAAA 60

Start of orf1

M N K E T V S I I M P V Y N 120
TATATAAGAAATTCTCAAATGAACAAAGAAACCGTTCAATAATTATGCCGTTACAAT

G A K T I I S S V E S I I H Q S Y Q D F 180
GGGGCCAAAACTATAATCTCATCAGTAGAATCAATTACATCAATCTTATCAAGATTTT

V L Y I I D D C S T D D T F S L I N S R 240
GTTTGATATCATTGACGATTGTAGCACCAGTACATTTCAACAGTCGA

Y K N N Q K I R I L R N K T N L G V A E 300
TACAAAACAATCAGAAAATAAGAATATTGCGTAACAAGACAAATTAGGTGTTGCAGAA

S R N Y G I E M A T G K Y I S F C D A D 360
AGTCGAAATTATGGAATAGAAATGGCCACGGGAAATATATTCTTTGTGATGCGGAT

D L W H E K K L E R Q I E V L N N E C V 420
GATTGTCGGCACGAGAAAAATTAGAGCGTAAATCGAAGTGTAAATAATGAATGTGTA

D V V C S N Y Y V I D N N R N I V G E V 480
GATGTTGATGTTCTAATTATTATGTTAGATAACAATAGAAATTGTTGGCGAAGTT

N A P H V I N Y R K M L M K N Y I G N L 540
AATGCTCCTCATGTGATAAAATTAGAAAATGCTCATGAAAACTACATAGGAATTG

T G I Y N A N K L G K F Y Q K K I G H E 600
ACAGGAATCTATAATGCCAACAAATTGGTAAGTTTATCAAAAAAGATTGGTCACGAG

D Y L M W L E I I N K T N G A I C I Q D 660
GATTATTGATGTCGGCTGGAAATAATTAAACAAATTGGTCTATTGTATTCAAGAT

N L A Y Y M R S N N S L S G N K I K A A 720
AATCTGGCGTATTACATGCGTTCAAATAATTCACTATCGGTAATAAAATTAAAGCTGCA

K W T W S I Y R E H L H L S F P K T L Y 780
AAATGGACATGGAGTATATAGAGAACATTACATTGTCCTTCAAAACATTATAT

Y F L L Y A S N G V M K K I T H S L L R 840
TATTTTTATTATATGCTCAAATTGGAGTCATGAAAAAAATAACACATTCACTATTAAAGG

Start of orf2, End of orf1

R K E T K K * 900
V K S A A K L I F L F L F T
AGAAAGGAGACTAAAAGTGAAGTCAGCGGCTAAGTTGATTTTTATTCCCTATTACAC

L Y S L Q L Y G V I I D D R I T N F D T 960
TTTATAGTCTCCAGTTGTATGGGTTATCATAGATGATCGTATAACAAATTGATACAA

K V L T S I I I F Q I F F V L L F Y L 1020
AGGTATTAACTAGTATTATAATTATTCAGATTTTTTGTTTATTATTTATCTAA

T I I N E R K Q Q K K F I V N W E L K L 1080
CGATTATAAAATGAAAGAAAACAGCAGAAAAATTATCGTGAACGGGAGCTAAAGTTAA

I L V F L F V T I E I A A A V V L F L K E 1140
TACTCGTTTCTTTGTGACTATAGAAATTGCTGCTGTAGTTATTCTAAAGAAG

G I P I F D D D P G G A K L R I A E G N 1200
GTATTCCTATATTGATGATGATCCAGGGGGCTAAACTTAGAATAGCTGAAGGTAATG

G L Y I R Y I K Y F G N I V V F A L I I 1260
 GACTTACATTAGATATTAAGTATTTGGTAATATAGTTGTGTTGCATTAATTATTC
 L Y D E H K F K Q R T I I F V Y F T T I 1320
 TTTATGAGCATAAAATTCAAACAGAGGACCATCATTTGTATATTTACAACGATTG
 A L F G Y R S E L V L L I L Q Y I L I T 1380
 CTTTATTTGGTTATCGTTCTGAATTGGTGTGCTCATTCTTCATATATATTGATTACCA
 N I L S K D N R N P K I K R I I G Y F L 1440
 ATATCCTGTCAAAGGATAACCGTAATCCTAAAATAAAAGAATAATAGGGTATTTTAT
 L V G V V C S L F Y L S L G Q D G E Q N 1500
 TGGTAGGGTTGTATGCTCGTTATCTAAGTTAGGACAAGACGGAGAACAAAATG
 D S Y N N M L R I I N R L T I E Q V E G 1560
 ACTCATATAATAATATGTTAAGGATAATTAATAGGTTAACATAGAGCAAGTTGAAGGTG
 V P Y V V S E S I K N D F F P T P E L E 1620
 TTCCATATGTTGTTCTGAATCTATTAAGAACGATTCTTCGACACCAGAGTTAGAAA
 K E L K A I I N R I Q G I K H Q D L F Y 1680
 AGGAATTAAAAGCAATAATAATAGAACATACAGGAAATAAGCATCAAGACTTATTTATG
 G E R L H K Q V F G D M G A N F L S V T 1740
 GAGAACGGTTACATAACAGTATTGGAGACATGGGAGCAAATTTTATCAGTTACTA
 T Y G A E L L V F F G F L C V F I I P L 1800
 CGTATGGAGCAGAACTGTAGTTTTTTGGTTCTGTATTCCCTTTAG
 G I Y I P F Y L L K R M K K T H S S I N 1860
 GGATATATACCTTTATCTTTAAAGAGAACGAAACCCATAGCTCGATAAATT
 C A F Y S Y I I M I L L Q Y L V A G N A 1920
 GCGCATTCTATTACATATCATTATGATTGCAACTTAGTGGCTGGAAATGCAT
 S A F F F G P F L S V L I M C T P L I L 1980
 CGGCCTTCTTTGGTCCTTCTCCGTATTGATAATGTGTACTCCTCTGATCTTAT

 Start of orf3
 M K I S V I T V T Y 2040
 L H D T L K R L S R N E N I S Y N C D L
 TGCAATGACGTTAAAGAGATTATCACGAAATGAAATATCAGTTAACTGTGACTTAT

 End of orf2
 N N A E G L E K T L S S L S I L K I K P 2100
 * AATAATGCTGAAGGGTTAGAAAAACTTTAAGTAGTTATCAATTAAAGAACCT
 F E I I I V D G G S T D G T N R V I S R 2160
 TTTGAGATTATTATAGTTGATGGCGCTCTACAGATGAAACGAATCGTGTCTAGTGA
 F T S M N I T H V Y E K D E G I Y D A M 2220
 TTTACTAGTATGAATATTACACATGTTATGAAAAGATGAAGGGATATATGATGCGATG
 N K G R M L A K G D L I H Y L N A G D S 2280
 AATAAGGGCGAATGTTGCCAAAGGCGACTTAATACATTATTTAACGCCGGCGATAGC
 V I G D I Y K N I K E P C L I K V G L F 2340
 GTAATTGGAGATATATAAAATATCAAAGAGCCATGTTGATTAAGTTGGCCTTTTC
 E N D K L L G F S S I T H S N T G Y C H 2400
 GAAAATGATAAACTCTGGGATTTCTCTATAACCCATTCAAATACAGGGTATTGTCAT

Figure 8/2

21/58

Q G V I F P K N H S E Y D L R Y K I C A	2460
CAAGGGGTGATTTCCAAAGAAATCATCAGAATATGATCTAAGGTATAAAATATGTGCT	
D Y K L I Q E V F P E G L R S L S L I T	2520
GATTATAAGCTTATTCAAGAGGTGTTCTGAAGGGTTAAGATCTCTATCTTGATTACT	
S G Y V K Y D M G G V S S K K R I L R D	2580
TCGGGTTATGTAAAATATGATATGGGGAGTATCTCAAAAAAAAGAATTTAAGAGAT	
K E L A K I M F E K N K K N L I K F I P	2640
AAAGAGCTTGCCAAAATTATGTTGAAAAAAATAAAAAACCTTATTAAGTTATTCCA	
I S I I K I L F P E R L R R V L R K M Q	2700
ATTTCAATAATCAAAATTATCCCTGAACGTTAAGAAGAGTATTGCGGAAATGCAA	
Start of orf4 End of orf3	
Y I C L T L F F M K N S S P Y D N E *	
M I M N K I	2760
TATATTTGTCTAACCTTATTCTCATGAAGAATAGTCACCATATGATAATGAATAAAAT	
K K I L K F C T L K K Y D T S S A L G R	2820
CAAAAAAAACTTAAATTTCGACTTTAAAAATATGATACATCAAGTGCTTAGGTAG	
E Q E R Y R I I S L S V I S S L I S K I	2880
AGAACAGGAAAGGTACAGGATTATATCCTGTCTGTTATTCAGTTGATTAGTAAAAT	
L S L L S L I L T V S L T L P Y L G Q E	2940
ACTCTCACTACTTCTCTATATTAACTGTAAGTTAACCTTACCTTATTAGGACAAGA	
R F G V W M T I T S L G A A L T F L D L	3000
GAGATTGGTGTATGGATGACTATTACAGTCTGGTGCTGCTGACATTGGACTT	
G I G N A L T N R I A H S F A C G K N L	3060
AGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCAATTGGCTGTGGCAAAATTT	
K M S R Q I S G G L T L L A G L S F V I	3120
AAAGATGAGTCGGCAAATTAGTGGTGGCTCACTTGCTGGATTATCGTTGTCAT	
T A I C Y I T S G M I D W Q L V I K G I	3180
AACTGCAATATGCTATATTACTCTGGCATGATTGATTGCAACTAGTAATAAAAGGTAT	
N E N V Y A E L Q H S I K V F V I I F G	3240
AAACGAGAATGTGTATGCAGAGTTACAACACTCAATTAAAGTCTTGTAAATCATATTGG	
L G I Y S N G V Q K V Y M G I Q K A Y I	3300
ACTTGGAAATTATTCAATGGTGTGCAAAAGTTATATGGGAATACAAAAAGCCTATAT	
S N I V N A I F I L L S I I T L V I S S	3360
AAGTAATATTGTTAATGCCATATTATATTGTTATCTATTACTCTAGTAATATCGTC	
K L H A G L P V L I V S T L G I Q Y I S	3420
GAAACTACATGCGGGACTACCAGTTAATTGTCAGCACTCTGGTATTCAATACATATC	
G I Y L T I N L I I K R L I K F T K V N	3480
GGGAATCTATTAAACAATTATCTTATTATAAAAGCGATTAATAAAGTTACAAAAGTTAA	
I H A K R E A P Y L I L N G F F F F I L	3540
CATACATGCTAAAAGAGAAGCTCCATATTGATATTAAACGGTTTTCTTTTTATT	
Q L G T L A T W S G D N F I I S I T L G	3600
ACAGTTAGGCACTCTGGCAACATGGAGTGGTGATAACATTATAATCTATAACATTGGG	

Figure 8/3

V T Y V A V F S I T Q R L F Q I S T V P 3660
 TGTTACTTATGTTGCTGTTTTAGCATTACACAGAGATTATTCAAATATCTACGGTCCC
 L T I Y N I P L W A A Y A D A H A R N D 3720
 TCTTACGATTATAACATCCCGTTATGGGCTGCTTATGCAGATGCTCATGCACGCAATGA
 T Q F I K K T L R T S L K I V G I S S F 3780
 TACTCAATTATAAAAAAGACGCTCAGAACATCATTGAAAATAGTGGGTATTCATCATT
 L L A F I L V V F G S E V V N I W T E G 3840
 CTTATTGGCCTTCATATTAGTAGTGTTCGGTAGTGAAGTCGTTAATATTGGACAGAAGG
 K I Q V P R T F I I A Y A L W S V I D A 3900
 AAAGATTCAAGGTACCTCGAACATTCAATAGCTTATGCTTATGGCTGTTATGATGC
 F S N T F A S F L N G L N I V K Q Q M L 3960
 TTTTCGAATACATTGCAAGCTTTAAATGGTTGAACATAGTTAAACAACAAATGCT
 A V V T L I L I A I P A K Y I I V S H F 4020
 TGCTGTTGTAACATTGATATTGATCGAACATTCCAGCAAAATACATCATAGTTAGCCATT
 G L T V M L Y C F I F I Y I V N Y F I W 4080
 TGGGTTAACTGTTATGTTGTAATGTTATATATATTGTAATTACTTTATATG

Start of orf5, End of orf4
M K M

Y K C S F K K H I D R Q L N I R G * 4140
 GTATAAAATGTAAGTTAAAAACATATCGATAGACAGTTAAATATAAGAGGATGAAAATG
 K Y I P V Y Q P S L T G K E K E Y V N E 4200
 AAATATATACCAAGTTACCAACCGTCAATTGACAGGAAAGAAAAGAAATATGTAATGAA
 C L D S T W I S S K G N Y I Q K F E N K 4260
 TGTCCTGGACTCAACGTTGAGATTTCATCAAAAGGAAACTATATTGAGAAGTTGAAATAAA
 F A E Q N H V Q Y A T T V S N G T V A L 4320
 TTTGGGAAACAAACCATGTGCAATATGCAACTACTGTAACTATGGAACGGTTGCTCTT
 H L A L L A L G I S E G D E V I V P T L 4380
 CATTTAGCTTCTTACGGTTAGGTATATCGGAAGGAGATGAAATTGTTCAACACTG
 T Y I A S V N A I K Y T G A T P I F V D 4440
 ACATATATAGCATCAATTGCTATAAAACACAGGAGECACCCCCATTTCGTTGAT
 S D N E T W Q M S V S D I E Q K I T N K 4500
 TCAGATAATGAAACTTGGCAATGTTACTGACATAGAACAAAAACTCAATAAA
 T K A I M C V H L Y G H P C D M E Q I V 4560
 ACTAAACCTATTATGTTGTCCTTATACGGACATECCATGTGATATGGAACAAATTGTA
 E L A K S R N L F V I E D C A E A F G S 4620
 GAATGCGGAAAGTAGAAATTGTTGTAATTGAAAGATTGCGCTGAAACCTTGGTTCT
 K Y K G K Y V G T F G D I S T F S F F G 4680
 AAATATAAGCTAAATATGTTGCAACATTGAGATATTCTACTTTTACCTTTTGGAA
 N K T I T T G E G G M V V T N D K T L Y 4740
 AATAAAACCTATTACTACAGGTGAAAGCTGGATGGTTGTCAGGAATGACAAAACACTTTAT
 D R C L H F K G Q G L A V H R Q Y W H D 4800
 GACCGCTGTTACATTAAAGCCAGGATTAGCTGACATAGGCAATATGCGCATGAC
 V I G Y N Y R M T N I C A A I G L A Q L 4860
 GTTATAGGCTACATTAGGATGACAATATCTGGCTGCTATAGGATTAGCGGCTTA

Figure 8/4

E Q A D D F I S R K R E I A D I Y K K N GAAACAAGCTGATGATTATACAGGAAACGTGAAATTGCTGATATTATAAAGGAAAT	4920
I N S L V Q V H K E S K D V F H T Y W M ATCAACAGTCTTGTACAAGTCACAGGAAAGTAAAGATGTTTTTCACACATTATTGGATG	4980
V S I L T R T A E E R E E L R N H L A D GTETCAATTCTAAGTACAGGACGGCAGAGGAAAGAGAGGAATTAGGAATCACCTTGCAGAT	5040
K L I E T R P V F Y P V H T M P M Y S E AAACTCATCGAAAGAGGCCAGTTTTTACCCGTCTTACCAACAGATGCCATTACTCGAA-	5100
K Y Q K H P I A E D L G W R G I N L P S AAATATCAAAGCACCTATAGCTGAGGATCTGGTTGGCCTGGAATTAAATTACCTAGT	5160
F P S L S N E Q V I Y I C E S I N E F Y TTECCCCAGCCTATCGAATGAGCAAGTTATTATATTGTGAATCTATTACGAATTATTAT	5220
End of orf5	Start of orf6
S D K *	M K I A L N S D
AGTGATAAAATAGCCTAAAATATTGTAAGGTCAATTGATGAAATTGCGTTGAATTGAGAT	5280
G F Y E W G G G I D F I K Y I L S I L E GGATTTTACGAGTGGGGGGGGTGGAAATTGATTTATTAAATATTCTGTCAATATTAGAA	5340
T K P E I C I D I L L P R N D I H S L I ACGAAACCAGAAATATGTATCGATATTCTTTACCGAGAAATGATATAACATTCTCTATA	5400
R E K A F P F K S I L K A I L K R E R P AGAGAAAAAGCATTCTTTAAAAGTATATTAAAGCAATTAAAGAGGGAAAGGCCT	5460
R W I S L N R F N E Q Y Y R D A F T Q N CGATGGATTCATTAAATAGATTAAATGAGCAACTATAGAGATGCCTTACACAAAAT	5520
N I E T N L T F I K S K S S A F Y S Y F AATATAGAGACGAATCTTACCTTATAAAAGTAAGAGCTCTGCCCTTTATTCATATTAA	5580
D S S D C D V I L P C M R V P S G N L N GATAGTAGCGATTGTGATGTTATTCTCCTGCGATGCGTGTCCCTCGGGAAATTGAAAT	5640
K K A W I G Y I Y D F Q H C Y Y P S F F AAAAAAGCATGGATTGGTTATATTATGACTTTCAACACTGTTACTATCCTTCATTTTT	5700
S K R E I D Q R N V F F K L M L N C A N AGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTAAATTGATGCTCAATTGCGCTAAC	5760
N I I V N A H S V I T D A N K Y V G N Y AATATTATTGTTAATGCACATTCAAGTTATTACCGATGCAAATAATATGTTGGAAATTAT	5820
S A K L H S L P F S P C P Q L K W F A D TCTGCAAAACTACATTCTCTCCATTAGTCCATGCCCTCAATTAAATGGTCCGCTGAT	5880
Y S G N I A K Y N I D K D Y F I I C N Q TACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTATTTATAATTGCAATCAA	5940
F W K H K D H A T A F R A F K I Y T E Y TTTGGAAACATAAGATCATGCAACTGCTTTAGGGCATTAAAATTATACTGAATAT	6000
N P D V Y L V C T G A T Q D Y R F P G Y AATCCTGATGTTATTAGTATGCAACGGGAGCTACTCAAGATTATGATTCCCTGGATAT	6060
F N E L M V L A K K L G I E S K I K I L TTTAATGAATGATGGTTTGGCAAAAAGCTCGGAATTGAATCGAAAATTAAAGATATTAA	6120

Figure 8/5

G H I P K L E Q I E L I K N C I A V I Q
GGGCATATACTAAACTGAAACAAATTGAATTAATCAAAATTGCATTGCTGTAATACAA 6180

P T L F E G G P G G V T F D A I A L G
CCAACCTTATTGAAAGCGGGCTGGAGGGGGTAACATTGACGCTATTGCATTAGGG 6240

K K V I L S D I D V N K E V N C G D V Y
AAAAAAAGTTATACTATCTGACATAGATGTCATAAAGAAGTTAATTGCGGTGATGTATAT 6300

F F Q A K N H Y S L N D A M V K A D E S
TTCTTCAGGCAAAAACATTATTCAATTAAATGACGCGATGGTAAAAGCTGATGAATCT 6360

K I F Y E P T T L I E L G L K R R N A C
AAAATTTTATGAACTACAACCTCTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGT 6420

End of orf6

A D F L L D V V K Q E I E S R S *
GCAGATTTCTTTAGATGTTGAAACAAGAAATTGAATCCGATCT TAATATATTCAA 6480

Start of orf7

M T K V A L I T G V T G Q D G S Y
GAGGTATATAATGACTAAAGTCGCTCTTATTACAGGTGTACTGGACAAAGATGGATCTTA 6540

L A E F L L D K G Y E V H G I K R R A S
TCTAGCTGAGTTTGCTTGATAAAGGGTATGAAGTTCATGGTATCAAACGCCAGCCTC 6600

S F N T E R I D H I Y Q D P H G S N P N
ATCTTTAAATACAGAACGCATAGACCATATTATCAAGATCCACATGGTCTAACCCAAA 6660

F H L H Y G D L T D S S N L T R I L K E
TTTCACTTGCACATGGAGATCTGACTGATTCTAACCTCACTAGAATTCTAACAGGA 6720

V Q P D E V Y N L A A M S H V A V S F E
GGTACAGCCAGATGAAGTATATAATTAGCTGCTATGAGTCACGTAGCAGTTCTTTGA 6780

S P E Y T A D V D A I G T L R L L E A I
GTCTCCAGAATATACAGCCGATGTCGATGCAATTGGTACATTACGTTACTGGAAGCAAT 6840

R F L G L E N K T R F Y Q A S T S E L Y
TCGCTTTAGGATTGGAAAACAAACGCGTTCTATCAAGCTAACCTCAGAATTATA 6900

G L V Q E I P Q K E S T P F Y P R S P Y
TGGACTTGTTCAGGAAATCCCTCAAAAGAACCTCAGCCTTTATCCTCGTCCCCCTTA 6960

A V A K L Y A Y W I T V N Y R E S Y G I
TGCAGTTGCAAAACTTACGCATATTGGATCACGGTAAATTATCGAGAGTCATATGGTAT 7020

Y A C N G I L F N H E S P R R G E T F V
TTATGCATGTAATGGTATATTGTCATGAATCTCACGCCGGAGAAACGTTGT 7080

T R K I T R G L A N I A Q G L E S C L Y
AACAGGAAAATTACTCGAGGACTTGCAAATATTGCACAGGCTTGGAAATCATGTTGTA 7140

L G N M D S L R D W G H A K D Y V R M Q
TTAGGAAATGGATTGTTACGGAGATTGGGACATGCAAAAGATTATGTTAGAATGCA 7200

W L M L Q Q E Q P E D F V I A T G V Q Y
ATGGTTGATGTTACAACAGGAGCAACCGAAGATTGTTGTGATTGCAACAGGAGTCCAATA 7260

S V R Q F V E M A A A Q L G I K M S F V
CTCAGTCCGTCAGTTGTCGAAATGGCAGCAGCACAACTGGTATTAAGATGAGCTTGT 7320

G K G I E E K G I V D S V E G Q D A P G	7380
TGGTAAAGGAATCGAAGAAAAAGGCATTGTAGATTCGGTTGAAGGACAGGATGCTCCAGG	
V K P G D V I V A V D P R Y F R P A E V	7440
TGTGAAACCAGGTGATGTCATTGTTGCTGTTGATCCTCGTTATTCCGACCAGCTGAAGT	
D T L L G D P S K A N L K L G W R P E I	7500
TGATACTTTGCTGGAGATCCGAGCAAAGCTAATCTCAAACCTGGTTGGAGACCAGAAAT	
T L A E M I S E M V A K D L E A A K K H	7560
TACTCTTGCTGAAATGATTTCTGAAATGGTTGCCAAAGATCTGAAGCCGCTAAAAAACAA	

Start of orf8, End of orf7	
M M M N K	
S L L K S H G F S V S L A L E *	7620
TTCTCTTTAAAATCGATGGTTTTCTGTAAGCTTAGCTCTGGAATGATGATGAATAAG	
Q R I F I A G H Q G M V G S A I T R R L	7680
CAACGTATTTTATTGCTGGTCACCAAGGAATGGTTGGATCAGCTATTACCCGACGCCCTC	
K Q R D D V E L V L R T R D E L N L L D	7740
AAACAACGTGATGATGTTGAGTTGGTTTACGTACTCGGGATGAATTGAACCTGTTGGAT	
S S A V L D F F S S Q K I D Q V Y L A A	7800
AGTAGCGCTGTTGGATTTTTCTTCACAGAAAATCGACCAGGTTATTGGCAGCA	
A K V G G I L A N S S Y P A D F I Y E N	7860
GCAAAAGTCGGAGGTATTTAGCTAACAGTTCTATCCTGCCATTATATGAGAAT	
I M I E A N V I H A A H K N N V N K L L	7920
ATAATGATAGAGGCGAATGTCATTGCTGCCACAAAATAATGAAATAACTGCTT	
F L G S S C I Y P K L A H Q P I M E D E	7980
TTCCCTCGGTTCGTCGTGTTACCTAAGTTAGCACACCAACCGATTATGAAAGACGAA	
L L Q G K L E P T N E P Y A I A K I A G	8040
TTATTACAAGGGAAACTTGAGCCAACAAATGAACCTTATGCTATCGAAAAATTGCAGGT	
I K L C E S Y N R Q F G R D Y R S V M P	8100
ATTAAATTATGTGAATCTTATAACCGTCAGTTGGCGTGATTACCGTTCAAGTAATGCCA	
T N L Y G P N D N F H P S N S H V I P A	8160
ACCAATCTTATGGTCCAAATGACAATTTCATCCAAGTAATTCTCATGTGATTCCGGCG	
L L R R F H D A V E N N S P N V V V W G	8220
CTTTGCGCCGTTTCATGATGCTGTGGAAAACAATTCTCGAATGTTGTTGGGGGA	
S G T P K R E F L H V D D M A S A S I Y	8280
AGTGGTACTCCAAAGCGTGAATTCTTACATGTAGATGATATGGCTCTGCAAGCATTAT	
V M E M P Y D I W Q K N T K V M L S H I	8340
GTCATGGAGATGCCATACGATATATGGCAAAAAAAACTAAAGTAATGTTGTCTCATATC	
N I G T G I D C T I C E L A E T I A K V	8400
AATATTGGAACAGGTATTGACTGCACGATTGTGAGCTTGCAGAACATAGCAAAAGTT	
V G Y K G H I T F D T T K P D G A P R K	8460
GTAGGTTATAAAGGGCATATTACGTTCGATACAACAAAGCCGATGGAGCCCTCGAAAA	
L L D V T L L H Q L G W N H K I T L H K	8520
CTACTTGATGTAACGCTTCTTCATCAACTAGGTTGGAATCATAAAATTACCCTTCACAAG	

Figure 8/7

End of orf8

G L E N T Y N W F L E N Q L Q Y R G *
 GGTCTGAAAATACATAACTGGTTCTTGAAAACCAACTTCATATCGGGGG TAATAA 8580

Start of orf9
 M F L H S Q D F A T I V R S T P L I S I
TGTTTTACATTCCAAGACTTGCCACAATTGTAAGGTCTACTCCTCTATTCTATAG 8640

D L I V E N E F G E I L L G K R I N R P
 ATTTGATTGTGGAAAACGAGTTGGCGAAATTGCTAGGAAAACGAATCAACGCCGG 8700

A Q G Y W F V P G G R V L K D E K L Q T
 CACAGGGCTATTGGTCGTTCTGGTAGGGTGTGAAAGATGAAAATTGCAGACAG 8760

A F E R L T E I E L G I R L P L S V G K
 CCTTGAAACGATTGACAGAAATTGAACTAGGAATTGCTTCGCTCTCTGTGGTAAGT 8820

F Y G I W Q H F Y E D N S M G G D F S T
 TTTATGGTATCTGGCAGCACTCTACGAAGACAATAGTATGGGGGGAGACTTTCAACGC 8880

H Y I V I A F L L K L Q P N I L K L P K
 ATTATATAGTTATAGCATCCCTCTAAATTACAACAAACATTGAAATTACCGAAGT 8940

S Q H N A Y C W L S R A K L I N D D D V
 CACAACATAATGCTTATTGCTGGCTATCGCAGCAAAGCTGATAAAATGATGACGATGTGC 9000

H Y N C R A Y F N N K T N D A I G L D N
 ATTATAATTGTCGCGCATATTTAACAAATGATGCGATTGGCTTAGATAATA 9060

Start of orf10 End of orf9

M S D A P I I A V V M A G G T G S
 K D I I C L M R Q *
AGGATATAATATGCTGATGCGCCAATAATTGCTGTAGTTATGGCCGGTGGTACAGGCAG 9120

R L W P L S R E L Y P K Q F L Q L S G D
 TCGCTTTGGCCACTTCTCGTAACATATCAAAGCAGTTTACAACCTCTGGTGA 9180

N T L L Q T T L L R L S G L S C Q K P L
 TAACACCTTGTACAAACGACTTTGCTACGACTTCAGGCCTATCATGTCAAAACCATT 9240

V I T N E Q H R F V V A E Q L R E I N K
 AGTGATAACAAATGAACAGCATCGCTTGTGGCTGAACAGTTAAGGGAAATAAATAA 9300

L N G N I I L E P C G R N T A P A I A I
 ATTAAATGGTAATATTCTAGAACCATGCGGGCAAATCTGCAACCAGCAATAGCGAT 9360

S A F H A L K R N P Q E D P L L L V L A
 ATCTGCGTTTCATGCGTTAACAGTAATCCTCAGGAAGATCCATTGCTTCTAGTTCTGC 9420

A D H V I A K E S V F C D A I K N A T P
 GGCAGACCACGTTAGCTAAAGAAAGTGTCTGTGATGCTATTAAATGCAACTCC 9480

I A N Q G K I V T F G I I P E Y A E T G
 CATCGCTAACAGGTAAAATTGTAACGTTGGAATTACCAAGAATATGCTGAAACTGG 9540

Y G Y I E R G E L S V P L Q G H E N T G
 TTATGGGTATATTGAGAGAGGTGAACATCTGTACCGCTTCAAGGGCATGAAAATACTGG 9600

F Y Y V N K F V E K P N R E T A E L Y M
 TTTTATTATGTAAGTTGCGAAAAGCCTAACGTGAAACCGCAGAATTGTATAT 9660

T S G N H Y W N S G I F M F K A S V Y L
 GACTTCTGGTAATCACTATTGGAATAGTGGAAATTGATGTTAAGGCATCTGTTATCT 9720

Figure 8/8

E E L R K F R P D I Y N V C E Q V A S S	9780
TGAGGAATTGAGAAAATTAGACCTGACATTACAATGTTGTGAACAGGTTGCCTCATC	
S Y I D L D F I R L S K E Q F Q D C P A	9840
CTCATACATTGATCTAGATTTATTGATTATCAAAAGAACAAATTCAAGATTGTCCTGC	
E S I D F A V M E K T E K C V V C P V D	9900
TGAATCTATTGATTTGCTGTAATGAAAAACAGAAAAATGTGTTGATGCCCTGTTGA	
I G W S D V G S W Q S L W D I S L K S K	9960
TATTGGTGGAGTGACGTTGGATCTTGGCAATCGTTATGGACATTAGTCTAAAATCGAA	
T G D V C K G D I L T Y D T K N N Y I Y	10020
AACAGGAGATGTAAAGGTGATATTAACCTATGATACTAAGAATAATTATATCTA	
S E S A L V A A I G I E D M V I V Q T K	10080
CTCTGAGTCAGCGTTGGTAGCCGCCATTGGAATTGAAGATATGGTTATCGTGCAAACCTAA	
D A V L V S K K S D V Q H V K K I V E M	10140
AGATGCCGTTCTGTCTAAAAAGAGTGATGTACAGCATGTAAAAAAATAGTCGAAAT	
L K L Q Q R T E Y I S H R E V F R P W G	10200
GCTTAAATTGCAGCAACGTACAGAGTATATTAGTCATCGTGAAGTTTCCGACCATGGGG	
K F D S I D Q G E R Y K V K K I I V K P	10260
AAAATTGATTGATTGACCAAGGTGAGCGATACAAAGTCAAGAAAATTATTGTGAAACC	
G E G L S L R M H H H R S E H W I V L S	10320
TGGTAGGGCTTCTTAAGGATGCATACCACATCGTTCTGAACATTGGATCGTGTCTTC	
G T A K V T L G D K T K L V T A N E S I	10380
TGGTACAGCAAAAGTAACCCCTGGCGATAAAACTAACTAGTCACCGCAAATGAATCGAT	
Y I P L G A A Y S L E N P G I I P L N L	10440
ATACATTCCCTGGCGCAGCGTATAGCTTGAGAATCCGGCATAATCCCTCTTAATCT	
I E V S S G D Y L G E D D I I R Q K E R	10500
TATTGAAGTCAGTCAGGGATTATTGGGAGAGGATATTATAAGACAGAAAGAACG	
End of orf10 Start of orf11	
Y K H E D * M K S L T C F K A Y D I R	10560
TTACAAACATGAAGATTAACATATGAAATCTTAACCTGCTTAAAGCCTATGATATTG	
G K L G E E L N E D I A W R I G R A Y G	10620
CGGGAAATTAGGCGAAGAACTGAAGATATTGCCTGGCGCATTGGCGTGCCTATGG	
E F L K P K T I V L G G D V R L T S E A	10680
CGAATTCTCAAACCGAAAACCATTGTTAGGCGGTGATGTCCGCCTCACCAAGCGAACG	
L K L A L A K G L Q D A G V D V L D I G	10740
GTAAAACCTGGCGCTTGCAGGTTACAGGATGCGGGCGTCGATGTGCTGGATATCGG	
M S G T E E I Y F A T F H L G V D G G I	10800
TATGTCCGGCACCGAACGAGAGATCTATTGCCACGTTCCATCTCGAGTGGATGGCGGCAT	
E V T A S H N P M D Y N G M K L V R E G	10860
CGAAGTTACCGCCAGCCATAACCCGATGGATTACAACGGCATGAAGCTGGTGCAGAACGG	
A R P I S G D T G L R D V Q R L A E A N	10920
GGCTCGCCCGATCAGCGGTGATACCGGACTGCGCGATGTCCAGCGTCTGGCAGAACCAA	
D F P P V D E T K R G R Y Q Q I N L R D	10980
TGACTTCCCTCCTGTCGATGAAACCAAACGTGGTCGCTATCAGCAAATCAATCTGCGTGA	

Figure 8/9

A Y V D H L F G Y I N V K N L T P L K L
 CGCTTACGTTGATCACCTGTTGGTTATATCAACGTAAAAACCTCACGCCGCTCAAGCT 11040
 V I N S G N G A A G P V V D A I E A R F
 GGTGATCAACTCCGGAACGGCGCAGCGGGTCCGGTGGTGGACGCCATTGAAGCCGATT 11100
 K A L G A P V E L I K V H N T P D G N F
 TAAAGCCCTCGCGCACCGGTGGAATTAAATCAAAGTACACAACACGCCGGACGGCAATT 11160
 P N G I P N P L L P E C R D D T R N A V
 CCCCAACGGTATTCTAACCCGCTGCTGCCGAATGCCGCACGACACCCGTAATGCGGT 11220
 I K H G A D M G I A F D G D F D R C F L
 CATCAAACACGGCGCGGATATGGGCATTGCGCTTTGATGGCGATTTGACCGCTTTCCCT 11280
 F D E K G Q F I E G Y Y I V G L L A E A
 GTTGTGACGAAAAAGGGCAGTTATCGAGGGCTACTACATGTGCGCTGGCAGAAGC 11340
 F L E K N P G A K I I H D P R L S W N T
 GTTCCCTCGAAAAAAATCCCGCGCGAAGATCATCCACGATCCACGTCTCTGGAACAC 11400
 V D V V T A A G G T P V M S K T G H A F
 CGTTGATGTGGTACTGCCGCAGGGCAGCCGGTAATGTCGAAAACGGACACGCCCT 11460
 I K E R M R K E D A I Y G G E M S A H H
 TATTAAAGAACGTATGCGCAAGGAAGACGCCATCTACGGTGGCGAAATGAGCGCTCACCA 11520
 Y F R D F A Y C D S G M I P W L L V A E
 TTACTTCCGTGATTCGCTTACTGCGACAGCGGATGATCCCGTGGCTGGTCGCCGA 11580
 L V C L K G K T L G E M V R D R M A A F
 ACTGGTGTGCCGTGAAAGGAAAAACGCTGGCGAAATGGTGCACGCCGGATGGCGCGTT 11640
 P A S G E I N S K L A Q P V E A I N R V
 TCCGGCAAGCGGTGAGATCAACAGCAAATGGCGCAACCGTTGAGGCAATTATCGCGT 11700
 E Q H F S R E A L A V D R T D G I S M T
 GGAACAGCATTAGCCCGAGGCCTGGATCGCACCGATGGCATCAGCATGAC 11760
 F A D W R F N L R S S N T E P V V R L N
 CTTTGCCGACTGGCGCTTTAACCTGCGCTCCACACCGAACCGGTGGTGCCTGGTTGAA 11820
 V E S R G D V K L M E K K T K A L L K L
 TGTGGAATCACCGGTGATGTAAGCTAATGGAAAAGAAAATGACTAAAGCTCTTAAATT 11880
End of orf11
 L S E *
 GCTAAAGTGAGTGATTATTCATTAATCATTAAGCGTATTAAGATTATTAAGTAAT 11940
 GTTATTGCCGTATATGATGAATATGTGGCTTTTATGTATAACGACTATACCGCAACT 12000

Start of H-repeat
 TTATCTAGGAAAAGATTAATAGAAATAAGTTTGACTGACCAATTGCATTCACGTC 12060
 ACGATTGAGACGTTCTTGTCTTAAGACATTTTCATCGTTATGTAATAACAAATGTG 12120
 CCTTATATAAAAGGAGAACAAATGGAACCTAAAATAATTGAGACAATAGATTTTATT 12180
 ATCCCTGTTACGATATTAGCCAAAGTTGTATCCTGCATCAGCCTGCAATATTCAC 12240
 GAGTGCTTGTAACTGAATACATGTCTGCCATTCCAGATGATAACGACGTCACTGCA 12300
 ATTGATGGTAAACACTCGGCACACTTATGACAAGAGTCGTCAGAGGAGTGGTTCAT 12360

GTCATTAGTGCCTTCAGCAATGCACAGTCTGGCCTCGGATAGATCAAGACGGATGAGA 12420
 AACCTAACGCGTTCACAGTTATTGATGAACTTCTAAATGATGGGTATTAAAGGAAAAA 12480
 TAATCATAACTGATGCGATGGCTGCCAGAAAGATATTGAGAGAAGATATAAAACAGA 12540
 GATGTGATTATTGCTGTAAAAGGAAATAAGAGTCGGCTTAATAGAGTCTTGAGG 12600
 AGATATTACGCTGAAAGAATTAAATAATCCAAAACATGACAGTTACGCAATTAGTAAA 12660
 AGAGGCACGGCAGAGACGATGTCGCTTCATATTGTTGAGATGCTCTGATGAGCTTA 12720
 TTGATTCACGTTGAATGAAAGGGCTGCAGAATTATGAATGGCAGTCCACTTCTCT 12780
 CAATAATAGCAGAGCAAAAGAAATCGAAATGACGATCAAATATTATATTAGATCTG 12840
 CTGCTTTAACCGCAGAGAAGTCGCCACAGTAATCGAAATCACTGGCGCATGGAGAATA 12900
 AGTTGCACAGTAGCCTGATGTGGTAATGAATGAAATCGACTATAATATAAGAAGGCGAGT 12960
 TGCATTCGAATGATTTCTAGAATGCGGCACATCGCTATTAAATATCTGACAATGATAATG 13020
 TATTCAAGGCAGGATTATCATGTAAGATGCGAAAAGCAGTCATGGACAGAAACTTCCTAG 13080
End of the H-repeat
 CGTCAGGCATTGCAGCGTGCAGGCTTCAATCTTGCAT TGGTTTGATAAGATATTTC 13140

Start of orf12
 M N L Y G I F G A G S Y G R E 13200
 TTTGGAGATGGAAAATGAATTGTATGGTATTGGTCTGGAAAGTTATGGTAGAGAA
 T I P I L N Q Q I K Q E C G S D Y A L V 13260
 ACAATAACCCATTCTAAATCAACAAATAAGCAAGAATGTGGTTCTGACTATGCTCTGGTT
 F V D D V L A G K K V N G F E V L S T N 13320
 TTTGTGGATGATGTTGGCAGGAAAGAAAGTTAATGGTTGAAGTGCTTCAACCAAC
 C F L K A P Y L K K Y F N V A I A N D K 13380
 TGTTTCTAAAGCCCTTATTAAAAAGTATTAAATGTTGCTATTGCTAATGATAAG
 I R Q R V S E S I L L H G V E P I T I K 13440
 ATACGACAGAGAGTGTGAGTCATATTACACGGGTTGAACCAATAACTATAAAA
 H P N S V V Y D H T M I G S G A I I S P 13500
 CATCCAAATAGCGTTGTTATGATCATACTATGATAGGTAGTGGCGCTATTATTCTCCC
 F V T I S T N T H I G R F F H A N I Y S 13560
 TTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTCATGCAAACATATACTCA
 Y V A H D C Q I G D Y V T F A P G A K C 13620
 TACGTTGCACATGATTGTCATAGGAGACTATGTTACATTGCTCCTGGGCTAAATGT
 N G Y V V I E D N A Y I G S G A V I K Q 13680
 ATGGATATGTTGTTATTGAAGACAATGCATATAGGCTCGGGTGCAGTAATTAAGCAG
 G V P N R P L I I G A G A I I G M G A V 13740
 GGTGTTCTAACGCCACTTATTGGCGGGAGCCATTAGGTATGGGGCTGTT
 V T K S V P A G I T V C G N P A R E M K 13800
 GTCACTAAAAGTGTCCCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAA
 R S P T S I * 13860
 AGATCGCCAACATCTATTAAATGGGAATGCGAAAACACGTTCAAATGGGACTAATGTT

30 / 58

AAAATATATATAATTCGCTAATTTACTAAATTATGGCTTCTTTAAGCTATCCTTAC 13920
TTAGTTATTACTGATAACAGCATGAAATTATAATACTCTGATAACATTATACGTTATT 13980
CAAGCCGCATATCTAGCGGTAACCCCTGACAGGAGTAAACAATG 14024

31/58

GTTGACAAATACCGACCGTATAATGAATCAAACGTTCTGGATTGGTATTTATCCAGGGCTT 60
 GACTACAGAGCATTAGATTATGTCGTAAGTAAGTTGAAGAATTGGTTAAATT 120
Start of abe
 M L D V N K K I L M T G A T 180
 CTAATTTAGGATAGGATGCTGATGTGAATAAGAAAATCCTAATGACTGGCGCTACTA
 S F V G T H L L H S L I K E G Y S I I A 240
 GCTTTGTAGGTACCCATCTACTACATAGTCATCAAAGGAAGGTTATAGTATTATTGCAT
 L K R P I T E P T I I N T L I E W L N I 300
 TAAAGCGTCCTATAACCGAGCCAACGATTATCAATAACCTTGATTGAATGGTTGAATATAC
 Q D I E K I C Q S S M N I H A I V H I A 360
 AAGATATAGAAAAAATATGTCAATCATCTATGAATATTGCGATTGTCCATATTGCAA
 T D Y G R N R T P I S E Q Y K C N V L L 420
 CAGACTATGGTCGAAACAGAACCCCTATATCTGAACAAATATAATGTAATGTCCTATTAC
 P T R L L E L M P A L K T K F F I S T D 480
 CAACAAGACTGCTTGAGTTAATGCCAGCGCTTAAACGAAATTCTTATTCTACTGACT
 S F F G K Y E K H Y G Y M R S Y M A S K 540
 CTTTTTTGGAAATATGAGAAGCACTATGGATATATGCGTCTTACATGGCATCTAAA
 R H F V E L S K I Y V E E H P D V C F I 600
 GACATTTGTAGAACTATCAAAATATACGTAGAGGAACATCCAGACGTTGTTTATAA
 N L R L E H V Y G E R D K A G K I I P Y 660
 ATTACGTTAGAACATGTTACGGTGAGAGGGATAAGCAGGTTAAATCCCGTATG
 V I K K M K N N E D I D C T I A R Q K R 720
 TTATCAAAAAAATGAAAAACAATGAAGATATTGATTGTACGATGCCAGGCAGAAAAGAG
 D F I Y I D D V V S A Y L K I L K E G F 780
 ATTTTATTATAGACGATGTTGGCTTCCGGCTATTGAAAATTAAAGGAGGGTTTA
 N A G H Y D V E V G T G K S I E L K E V 840
 ACGCTGGACACTATGATGTCGAGGTGGGACTGGAAAATCGATAGAGCTAAAGAAGTGT
 F E I I K K E T H S S S K I N Y G A V A 900
 TTGAGATAATAAGAACGCATAGTAGTAGTAAGATAATTGAGTCAGTTGCAGTTGCGA
 M R D D E I M E S H A N T S F L T R L G 960
 TGCGTGATGATGAGATTATGGAGTCACATGCAAATACCTTTCTTGACTCGATTAGGTT
End of abe Start of wzx
 M
 W S A E F S I E K G V K K M L S M K E * 1020
 GGAGTGCCGAGTTCTATTGAGAAGGGTGTGAAAAAAATGTGAGTATGAAAGAG TAAT
 N R I I R M L G V D K A I R Y V I F G K 1080
 GAATCGTATTATAGAATGTTAGGTGTAGATAAAGCAATTGTTATGTTATTTGGTAA
 I I S V L T G L L I M L I S H H L S K 1140
 GATAATATCTGTATTAACGGGTTACTGTTAATAATGTTAATATCACACCCATTATCTAA
 D A Q G Y Y Y T F N S V V A L Q I I F E 1200
 AGACGCACAGGGCTATTATACATTAAATTCAATTGAGTAGTGGCACTACAGATAATATTGA

Figure 9/1

L G L S T V I I Q F A S H E M S A L K Y	1260
ATTGGGGCTATCAACGGTAATCATTCAATTGCTAGCCATGAAATGTCAGCGTTAAAATA	
D Y S E R D I I G E S K N K Q R Y L S L	1320
TGATTATTCTGAACGAGATATTAGGTGAAAGTAAAATAAGCAACGTTACCTATCGTT	
F R L A I K W Y A V I A L L I I L I V G	1380
ATTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTGCTAATAATATTAATAGTCGG	
P I G Y V F F T Q K E G L G V P W Q G A	1440
TCCCATCGGGTATGTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTGGCAAGGGC	
W L L L T I V T A F N I F L V S V L S V	1500
ATGGTTATTATTAACAATAGTTACAGCTTTAATATTTCTTGTACTTCGT	
A E G S G L I T D V N K M R M Y Q S L L	1560
CGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCGCTGTT	
A G I L A V S L L I S G F G L Y A T S A	1620
AGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTGGACTATATGCTACGTCTGC	
I A I S G T I I F S I F S Y K Y F K K I	1680
AATAGCTATTTCAGGGACTATCATATTCTCCATATTTCATATAAGTATTAAAAAAAT	
F L Q S L K H K N K Y T E G G I S W V N	1740
TTTCCTGCAATCTTAAAGCATAAAATATACTGAAGGTGGTATTTCATGGGTTAA	
E I F P M Q W R I A L S W M S G Y F I Y	1800
TGAAATATTCCATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTATTAA	
F V M T P I A F K Y F G A I Y A G Q L G	1860
TTTGTTATGACCCCCATTGCATTCAAATATTCGGGCTATATATGCAGGGCAGTTAGG	
M S L T L C N M V M A T G L A W I S T K	1920
GATGTCTTAAACATTATGCAATATGGTAATGGCTACGGGCTGGTGGATATCCACTAA	
Y P K W G V M V S N K Q L A E L S K S F	1980
ATATCCAAAATGGGAGTAATGGTTCAACAAACAGCTGGAACTGAGTAAATCGTT	
K S A V M Q S S F F V L T G L T G V Y I	2040
CAAAAGTGCAGTAATGCAATCATCCTTTGTCTGACAGGATTAACGGTGTACAT	
S L W L L K L S G S N I G E R F L G L Q	2100
TTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTGTTGGGATTGCA	
D F F F L S L A I I G N H I V A C F A T	2160
GGATTTTCTTTATCTTAGCAATTATGGTAATCACATTGAGCTGCTTGCAC	
Y I R A H K T E K M T L A S C I M A L L	2220
CTATATAAGAGCGCATAAAAGTACGACATTGGCATCATGTATAATGGCTCTCTT	
T I T T M L F V A Y L E Y S R F Y M L M	2280
GACTATAACTACAATGGTTGTCATATTAGAGTACTCGAGGTTCTACATGTTAAT	
Y A A L T W L Y F V P Q T Y I I F K R F	
S L K D	
GTATGCAGCACTAACGTGGTATATTGTTCCCTAAACCTATATAATCTTAAAGATT	2340

Figure 9/2

Start of *wbaR* End of *wzx*

K S S Y E *		
M S K K P L L T I A I P T Y N R		
CAAGAGTTCTT <u>ATGAG</u> TAAAAACCTCTTACTATTGCTATTCCGACATATAACCGCT	2400	
S S C L A R L L D S I I Q Q E N Y C H D		
CTTCATGTTGGCTCGTTACTTGATAGTATAATTCAACAGGAGAACTATTGTCATGATG	2460	
E L E V I V C D N A S T D E T A R I A K		
AACTCGAGGTTATTGTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGCCAAGA	2520	
S G L D K I R N S T Y H L N E E N L G M		
GTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGGAATGG	2580	
D G N F Q K C F E L S N G K Y L W M I G		
ATGGTAACCTCCAGAAATGTTGAGTTATCAAATGGAAAATATCTTGGATGATTGGCG	2640	
D D D L I V K N G I S K V F S I L K S R		
ATGATGATCTAATAGTCAAAATGGTATTCGAAGGTTTTCGATATTAAAGTCCCGGC	2700	
P A L D M V Y V N S A A K T E L N Y N A		
CTGCATTAGATATGGTGTATGTAATTCAAGCAGACTGAGTTAAACTATAATGCTG	2760	
D V R T S F Y T N D V D F I S D V K V M		
ATGTGAGGACGTCATTCTACACAAATGATGTAGATTTCAGACGTGAAAGTTATGT	2820	
F T F I S G M I C K K T D A I V K A V G		
TCACGTTATTCGAATGATATGTAAGAAAATGATGCAATTGCAAAGCCGTTGGTA	2880	
I F S P Q T T G K Y L M H L T W Q L P L		
TTTCAGTCGCAAACACTGGAAAATATCTTATGCATTAAACATGGCAATTGCCATTAC	2940	
L K Q G G E F A V I H N N I I E A E P D		
TTAACACAGGGTGGAGAGTCGCAGTTATCCATAATAATTGAGGCTGAGCCAGATA	3000	
N S G G Y H L Y K V F S N N L A T I F D		
ATTCAAGGTGGATATCATTATATAAGGTTTTCTAATAATTGAGGCTGAGCCAGATA	3060	
V F Y P R E H R V S K R V R A S A C L F		
TTTTTATCCCAGAGAGCACCGTGTAAAGTAAAGAGTCGCATCAGCATGTTATTCT	3120	
L L N F I G D E D K T K N F A T N N Y L		
TACTTAACCTCATAGGCATGAAGATAAAACCAAAATTGCTACAAATAATTATTAA	3180	
R D C D S A F I D L I I Y K Y G L R F F		
GAGATTGCGATAGTCGATTTATAGATTATAATTATAATTGGCTTAGGTTTTCT	3240	
Y L Y P K T V P L F R K I K Y I I K T V		
ATCTATATCCTAAAATGTGCCTTATTAGAAAATAATTATAAAAGACGGTTT	3300	
End of <i>wbaR</i>		
L M R K *		
TAATGCGGAAAATAAAATTATTCAAGATGGTTGCTGAAAACGACTTATAGGACTATCTA	3360	
Start of <i>wbaL</i>		
M F V Y S L R L K L N L I I S L L S K V		
<u>ATGTTGTCTATAGTTAAGATAAAATTAAATCTTATCATATCATTATTGAGTAAAGTT</u>	3420	
R R K S K A K F L V L L S G Y D F K M V		
AGGCAGGAAATCAAAGCAAAGTTCTGCTTAGCGGATATGATTAAAATGGTT	3480	

Figure 9/3

G K N F K L N V K P Y S A K N N T S S K GGAAAGAATTAAATTGAATGTCAACCTTACTCTGCAAAAAATAACACCTCTTCCAAA	3540
W G S M R V G D N C W I E A V Y N Y G D TGGGGTAGTATGCGGGTTGGTATACTGCTGGATTGAAGCTGTATATAATTATGGTGAT	3600
E K F E P Y L Y I G D R I C L S D N V H GAAAAATTGAAACCTTATTTGTACATAGGTGATCGTATATGTTAAGTGATAATGTTCAT	3660
I S C V S C L I L E N D I L I G S K V Y ATTTCTTGCCTATCATGTTAATTTAGAAAACGATATATTAAATTGGTAGCAAAGTTAT	3720
I G D H S H G S Y K V C S P K I E P P A ATAGGCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCA	3780
N K P L G D I A P I K I G N C C W I G D AATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAAATTGCTGGATTGGAGAT	3840
N A V I L A G S E I C D G C V I A A N S AATGCAGTAATTCTGGCTGGTAGTGAAATTGTGATGGCTGTGTAATCGCAGCTAATTCA	3900
V V K D L K V D K P C L I G G V P A K V GTCGTCAAGGATTAAAGTCGATAAGCCATGTTAATTGGTGGGGTTCTGCTAAAGTA	3960
End of wbaL Start of wbaQ	
I K V F * M N V F I S I C I P S Y N R A ATAAAGGTATTTAAAATGAATGTTTTATCAGTATTGTATAACCGTCTATAATAGAGC	4020
E F L E P L L D S I Y N Q D Y C L K N N TGAGTTTTAGAGCCACTACTGGATAGCATATATAATCAAGATTATTGTTAAAGAATAA	4080
D F E V I V C E D K S P Q R D E I N S I TGATTTGAGGTATTGTTGTGAAAGATAATCTCACAGAGAGATGAGATAAACTCTAT	4140
I E N Y K A K N N K Q N L Y V N F N E D TATCGAAAATATAAGCAAAAAATAATAACAAATCTTATGTTAATTCAATGAAGA	4200
N L G Y D K N L K K C I S L T T G K Y C TAATTTAGGCTATGATAAGAATTAAAAATGCATTAGTTGACGACAGGTAAATATTG	4260
M I M G N D D L L A D G A L S K I V K V CATGATCATGGCAACGATGATCTATTAGCAGATGGAGCGTTATCAAAATAGTGAAGT	4320
L K A N P E I V L A T R A Y G W F K E N TTTGAAGGCTAACCTGAAATTGTATTGGCTACGGAGCGTATGGTGGTTAAGGAAA	4380
P N E L C D T V R H L T D D T L F Q P G TCCGAATGAGTTATGTGATACTGTCGTCATTAAACAGACGATACTTTATTTCAGCCGG	4440
A D A I K F F R R V G V I S G F I V N GGCTGATGCCATTAAATTCTTCCGTAGAGTTGGAGTTATTTCAGGCTTATTGTCAA	4500
A E K A K K L S S D L F D G R L Y Y Q M TGCTGAAAAGCAAAAAACTATCGAGTGATTTGATGGCGTTATATTCAAAT	4560
Y L A G M L M A E G Q G Y Y F S D V M T GTACCTTGCTGGTATGCTAATGGCTGAAGGTAGGGATACTATTAGCGACGTGATGAC	4620

Figure 9/4

L S R D T E A P D F G N A G T E K G V F	4680
ATTGTCGAGGGATACAGAGGCTCTGACTTGGTAACGCTGGAACTGAAAAAGGAGTTT	
T P G G Y K P E G R I H M V E G L L I I	4740
CACCCCCGGGGGGTATAAAACCAGAGGGCCGTATACATATGGTTGAAGGCTTGTGCTAAT	
A K Y I E D T T K I D G V Y A G I R K D	4800
TGCAAAATATAGAAGATAACAACAAAAATTGATGGCGTTATGCTGGAATTAGAAAAGA	
L A N Y F Y P Y I R D Q L D L P L Y T Y	4860
CTTAGCGAACTATTTTATCCTTATATTGAGATCAACTCGACTTGCCTTTACTTA	
I K M I N K F R K M G F S N E K L F Y V	4920
TATTAAAATGATAAATAAAATTGCGAAAATGGGATTTCAAATGAAAAGCTTTCTATGT	
H A F L G Y V L K R R G Y D A L I K Y I	4980
GCATGCCTTTAGGGTATGTAACAAACGGAGGGCTATGATGCTTAATTAAATACAT	
End of wbaQ	
R S K K G G T P R L G I *	5040
TCGTAGCAAAAAGCGGTACTCCCGCTTGGTATTAAACCTCCACTTCAAAAATGT	
TATGAATATACTTCTGCTGCGATATTAGGCCTAACCTATTCTCCATATATTAGTC	5100
Start of wzy	
M L P F P P G A I L R D V L N V	5160
GTGGATGGTGGGT <u>ATGCTGCCATTCCACCAGGAGCAATCCTAAGGGATGTACTCAATGT</u>	
F F V A L V L V R F V I D R K K T Y F P	5220
ATTTTTTGTGGCGTTAGTGCTAGTCGATTGTCATTGATAGGAAAAAAACTTATTCCC	
L V F T I F S W S A V I L W V I A L T I	5280
GTTGGTTTTACTATTTTCATGGTCGGCGTAATACTATGGTAATAGCGTTAACTAT	
F S P D K I Q A I M G G R S Y I L F P A	5340
ATTCTCACCGATAAAATTCAAGCAATTATGGGGGGCGGAGTTATATTATTCCC	
V F I A L V I L K V S Y P Q S L N I E K	5400
AGTTTCATAGCATTAGTGAATTAAAGTATCATACCCGCAATCCTAAATATTGAAAA	
I V C Y I I F L M F M V A T I S I I D V	5460
AATAGTTGCTACATAATTCTAATGTTATGGTTGCGACAATATCTATTATTGATGT	
L M N G E F I K L L G Y D E H Y A G E Q	5520
ACTAATGAATGGAGAGTTCATTAATTGCTCGGATATGATGAGCATTATGCAGGAGAAC	
L N L I N S Y D G M V R A T G G F S D A	5580
ATTAAACCTTAATTAATAGCTATGATGGATGGTCCGGCTACAGGCAGTTAGTGATGC	
L N F G Y M L T L G V L L C M E C F S Q	5640
TCTCAATTGGATATGCTCACATTAGGTGTTATGTATGGAGTGTGTTCCCA	
G Y K R L L M L I I S F V L F I A I C M	5700
AGGATATAAAAGATTATTGATGCTTATTAGTTGTGCTATTAGCGATCTGCAT	
S L T R G A I L V A A L I Y A L Y I I S	5760
GAGTCCTACTAGAGGAGCAACTTGTGCTGCGCTTATTACGCACCTTATATAATTTC	
N R K M L F C G I T L F V I I I P V L A	5820
AAATCGGAAGATGCTTTGTGGAATAACTTATTGTAATAATTACCCGTTAGC	

I S T N I F D N Y T E I L I G R F T D S	5880
AATTTCTACTAATATTTGACAACATAACAGAAATTGATCGGCAGGTTACAGATTC	
S Q A S R G S T Q G R I D M A I N S L N	5940
GTCTCAGGCATCGCGTGGATCTACACAGGGCGGATAGATATGGCAATTAAATTCAATTAAA	
F L S E H P S G I G L G T Q G S G N M L	6000
CTTCCTGTCAGAACATCCATCAGGTATAGGTCTGGTACTCAAGGTTCAGGAAACATGCT	
S V K D N R L N T D N Y F F W I A L E T	6060
TCGGTAAAAGATAATAGGTTAACACGGATAATTATTTCTGGATCGCCCTTGAGAC	
G I I G L I I N I I Y L A S Q F Y S S T	6120
TGGTATTATTGGCTTAATCATAAATATTATTCAGTCAATTCTTCAC	
L L N R I Y G S H C S N M H Y R L Y F L	6180
TTTACTAAATAGAATATGGCAGTCATTGTAGCAATATGCACTATAGATTATTTCT	
F G S I Y F I S A A L S S A P S S S T F	6240
CTTGGGAAGTATATTTATAAGTGCAGCGTTAAGTTCAGCACCTCGTCATCAACTTT	
S I Y Y W T V L A L I P F L K L T N R R	6300
TTCTATATATTATTGGACAGTTAGCTTGATTCCATTAAAATTAACAAATAGACG	
End of wzy Start of wbaW	
C T R * M N N K K V L M D I S W S N K G	6360
GTGCACGCGATAATGAATAATAAAAGGTTTGATGGATATTAGTTGGTCTAATAAAAGGG	
G I G R F T D E I S K L L C D I S K E E	6420
GGGATTGGACGTTTACTGATGAAATTCTAAACTACTATGTGATATATCTAAGGAGGAA	
L Y R K C A S P L A P L G L A V N I F L	6480
CTATATAGAAAATGTGCTCTCCGCTGGCCCCATTAGGTTAGCAGTCATAATTCTG	
R K K T D V V F L P G Y I P P L F C S K	6540
CGAAAGAAAATGATGTGGTTCTCCTGGCTATATTCCACCACTTTTGTTGCGAAA	
K F I I T I H D L N H L D L N D N S S L	6600
AAGTTCTATAATAACAATACATGATCTAAATCATCTGGATTAAATGATAATTCTCTCTT	
F K R L F Y N F I I K R G C R K A Y K I	6660
TTTAAGAGGTTATTTATAATTAAAGCGCGTTGTAGAAAAGCATATAAAATA	
F T V S N F S K E R I V A W S G V N P N	6720
TTTACAGTTCGAATTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCTAAT	
K I V T V Y N G V S S L F N A D V K P L	6780
AAAATAGTCACGGTATATAATGGGTATCTAGTCTATTAAATGCCATGTAAAACCATTG	
N L G Y K Y L L C V G N R K T H K N E K	6840
AATTTAGGCTATAAAATATTGCTATGTGTAGGAAACAGAAAACATCAAAGAATGAGAAG	
C V I S A F A K A D I D P S I K L V F T	6900
TGTGTTATATCTGCCTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTACT	
G N P C N D L E K L I I Q H G L S E R V	6960
GGTAATCCTTGTAAATGATTTAGAAAAACTAATAACACATGGTTAAGTGAACGTGTA	

K F F G F V S E K D L P S L Y K G S L G	7020
AAGTTCTTGGGTTCGTGTCTGAAAAAGATTACCATCGTTATAAAGGGCTCGTTAGGA	
L V F P S L Y E G F G L P V V E G M A C	7080
TTAGTTTCCCTTCTTATATGAAGGTTGGATTACCTGTAGTGGAGGGCATGGCCTGT	
G I P V L T S L T S S L P E V A G D A A	7140
GGTATTCCCTGTATTAACCTCTCACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCG	
I L V D P L S E D A I T K G I S R L I N	7200
ATTCTTGTGACCCCTTTGAAAGATGCTATTACTAAAGGAATTGAGGTTAATTAAAT	
D S E L R K H L I Q K G L L R A K R F N	7260
GATTCTGAACCTCGTAAGCATTAAATCCAAAAGGGCTTTGCGGGCAAAGAGGTTCAAT	
Start of wbaZ	
W Q N V V S E I E M V L T E A C D G N K	
M E I N	
TGGCAAAACGTGGTTAGTGAGATTGAAATGGTACTGACAGAGGCATGTG <u>ATGAAATAAA</u>	7320
End of wbaW	
*	
E I K I S L V H E W L L S Y A G S E Q V	7380
TGAAATAAAATATCTCTCGTTCATGAGTGGTTATTAGTTATGCAGGCTCCGAACAGGT	
S S A I L H V F P E A K L Y S V V D F L	7440
ATCATCTGCCATCCTGCATGTTTCTGAAGCGAAGTTATTCGGTGGTTGATTTCT	
T D E Q R R H F L G K Y A T T T F I Q N	7500
AACGGATGAAACAAAGAAGACATTCTGGGAAATATGCGACTACCACATTATTCAAAA	
L P K A K K F Y Q K Y L P L M P L A I E	7560
TTTACCTAAAGCTAAAAATTACAGAAAATTTACCACTAATGCCACTGGCTATTGA	
Q L D L S D A N I I I S S A H S V A K G	7620
ACAACTTGATTATCAGATGCTAATATCATCATTAGTAGCGCCCATCCGTTGCAAAGG	
V I S G P D Q L H I S Y V H S P I R Y A	7680
TGTTATTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCCTATTGATATGC	
W D L Q H Q Y L N E S N L N K G I K G W	7740
GTGGGATTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGAATTAAAGGTTG	
L A K W L L H K I R I W D S R T A N G V	7800
GTTAGCAAAATGGCTTCTTCACAAAATACGAATTGGGATTCTCGAACCGCAAATGGGT	
D H F I A N S Q Y I A R R I K K V Y R R	7860
TGATCATTCTAGCTAATCTCAATATCGCGCGTAGAATTAAAAAGTATACAGACG	
E A S V I Y P P V D V D N F E V K N E K	7920
TGAGGCTTCAGTTATATCCGCCTGTAGATGTGGATAATTGAAAGTAAAAATGAAAA	
Q D Y Y F T A S R M V P Y K R I D L I V	7980
GCAAGACTATTATTCACAGCATCCCGTATGGTACCCCTACAAACGTATTGATCTTATTGT	
E A F S K M P E K K L V V I G D G P E M	8040
CGAACGCTTGTAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGTTATCAATCTTCC	
K K I K S K A T D N I K L L G Y Q S F P	8100
GAAAAAAATAAGAGCAAGGCTACAGACAATATAAAATTGCTCGTTATCAATCTTCC	

Figure 9/7

V L K E Y M Q S A R A F V F A A E E D F	8160
TGTTTAAAGAGTATATGCAGAGGCCAGGGCGTTGTTTGCAAGCGGAAGAGGACTT	
G I I P V E A Q A C G T P V I A F G K G	8220
TGGAATAATACCTGTCGAAGCTCAAGCTGCGGTACCCCTGTTATTGCCTTGGGAAGGG	
G A L E T V R P L G V E E P T G I F F K	8280
TGGGGCCTAGAAACCGTCGCCACTAGGTGTAGAGGAACCGACTGGCATTCTCAA	
E Q N I A S L H E A V S E F E K N A S F	8340
GGAACAGAATATTGCTTCTTGATGAAGCTGTTAGTGAATTGAAAAAAATGCATCATT	
F T S Q A C R K N A E K F S R S R F E Q	8400
TTTTACATCTCAGGCTTGTAGAAAAAAATGCAGAAAAATTCTCGATCAAGATTGAACA	
E F K N F V N E K W N L F K T E Q I I K	8460
AGAATTTAAGAACTTTGTTAATGAAAAGTGAATCTTCAAAACAGAACAGATTATTAA	
End of wbaZ Start of manC	
M S K L I P V I M A G G I	
R *	8520
ACGTAAATTATGGTTATTGA <u>ATGTCTAAATTAATACAGTAATAATGGCCGGTGGGATT</u>	
G S R L W P L S R E E H P K Q F L S V D	8580
GGTAGCCGTTGTGCCACTTCACGTGAAGAGCATTCCGAAACAGTTTAAGCGTAGAT	
G E L S M L Q N T I K R L T P L L A G E	8640
GGTGAATTATCTATGCTGCAAAACACCATTAAAGATTGACTCCTCTTTGGCTGGAGAA	
P L V I C N D S H R F L V A E Q L R A I	8700
CCTTTAGTCATTTGTAATGATAGTCACCGCTTCCTGTCGCTGAACAACCTCGAGCTATA	
N K L A N N I I L E P V G R N T A P A I	8760
ATAAAACTAGCAAATAACATCATATTAGAGCCAGTGGGGCGTAATACAGCCCCAGCTATA	
A L A A F C S L Q N V V D E D P L L L V	8820
GCGCTGGCCGTTTGTTCACTTCAGAATGTCGTCGATGAAGACCCGCTTGCTGTC	
L A A D H V I R D E K V F L K A I N H A	8880
CTTGCTGCGGATCATGTCATCCCGCATGAGAAAGTGTAAAGCTATCAATCACGCT	
E F F A T Q G K L V T F G I V P T Q A E	8940
GAATTTTTGCAACACAAGGTAGTAACGTTGGTATTGTACCCACACAGGCCGAA	
T G Y G Y I C R G E A I G E D A F S V A	9000
ACTGGCTACGGTTATTTGAGAGGTAGCAAGCAATCGGGGAAGATGCTTTCTGTAGCC	
E F V E K P D F D T A R H Y V E S E K Y	9060
GAATTTGTAGAGAACGCTGATTCGATACAGCGCGTATTATGTAGAATCAGAGAAATAT	
Y W N S G M F L F R A S S S Y L Q E L K D	9120
TATTGGAACAGCGGTATGTTCTATTTCGTGCAAGTAGTTACTTACAAGAATTAAAGGAT	
L S P D I Y Q A C E N A V G S I N P D L	9180
CTGTCCCCGATATTACCAAGCATGTGAAAATGCGGTAGGGAGTATTAATCCTGATCTT	
D F I R I D K E A F A M C P S D S I D Y	9240
GATTTATCCGTATTGATAAAGAACGATTGCAATGTGCCCTAGTGATTCTATCGATTAT	

A V M E H T R H A V V V P M N A G W S D	
GCGGTAATGGAACATACTAGGCATGCAGTTGCGTACCGATGAATGCCGGCTGGTCAGAT	9300
V G S W S S L W D I S K K D P Q R N V L	
GTGGGGTCATGGTCTTCACTGTGGATATTCTAAGAAAGATCCACAACGTAATGTATTA	9360
H G D I F A Y N S K D N Y I Y S E K S F	
CATGGCGATATTTGCATATAATAGTAAAGATAATTATATCTATTCTGAAAAATCGTTT	9420
I S T I G V N N L V I V Q T A D A L L V	
ATTAGTACAATCGGAGTAAATAATTAGTTATCGTCAGACAGCAGATGCATTATTAGTA	9480
S D K D S V Q D V K K V V D Y L K A N N	
TCTGATAAAAGATTCACTGGAGATGTTAAAAAGTTGTTGATTATTTAAAAGCTAATAAT	9540
R N E H K K H L E V F R P W G K F S V I	
AGAAACGAACATAAAAACATTAGAGGTTTCCGACCGTGGGAAATTAGCGTAATT	9600
H S G D N Y L V K R I T V K P G A K F A	
CATAGTGGCGATAATTAGTTAAAGAATAACTGTTAAACCAGGCGCGAAGTTGCT	9660
A Q M H L H R A E H W I V V S G T A C I	
GCTCAGATGCATCTCCATCGTGTGAGCATTGGATAGTGGTATCTGGTACTGCTTGTATT	9720
T K G E E I F T I S E N E S T F I P A N	
ACTAAGGGGAAAGAAATTTCGAGAATGAATCACATTATACCTGCTAAT	9780
T V H T L K N P A T I P L E L I E I Q S	
ACAGTTCATACGTTAAAAACCCCGCAGTATTCCATTAGAACTAATAGAAATTCAATCT	9840
G T Y L A E D D I I R L E K H S G Y L E	
GGCACCTATCTGGAGGATATTTCGCTGGAGAAACATTCTGGATATCTGGAG	9900
End of manC Start of manB	
*	
M K N I Y N T Y D V I N K S G I N	
TAATGAATT <u>GATGAAAAA</u> TATATAACTTACGATGTTATCAACAAATCTGGAATTAA	9960
F G T S G A R G L V T D F T P E V C A R	
TTTTGGAACCAAGTGGTGCCTGGCGCCCTTGTACCGATTACCCGAAGTTGCGCAG	10020
F T I S F L T V M Q Q R F S F T T V A L	
ATTTACCATTCCTTTGACAGTAATGCAGCAAAGATTCTCATTTACAACGGTTGCGCT	10080
A I D N R P S S Y A M A Q A C A A A L Q	
CGCAATTGATAATCGCCAAGCAGTTACGCGATGGCTCAAGTGTGCGCTGCTTGC	10140
E K G I K T V Y Y G V I P T P A L A H Q	
AGAAAAGGAATTAAACCGTTACTATGGCGTAATTCAACACCTGCTTAGCTCATCA	10200
S I S D K V P A I M V T G S H I P F D R	
ATCAATTCCGATAAAAGTACCTGCAATCATGGTACTGGCAGTCATATCCCTTTGACCG	10260
N G L K F Y R P D G E I T K D D E N A I	
TAATGGCCTGAAATTATAGACCAGATGGTGAAATTACTAAAGATGATGAGAATGCTAT	10320
I H V D A S F M Q P K L E Q L T I S T I	
TATTGATGGTACTGCCTCATTATGCAGCCTAAGCTGAACAATTGACAATTCCACAAAT	10380

A A R N Y I L R Y T S L F P M P F L K N CGCTGCTAGAAATTATATTCTACGATATAACCTCATTATTCCAATGCCATTCTTGAAAAA	10440
K R I G I Y E H S S A G R D L Y K T L F TAAGCGCATTGGAATTATGAGCATTCTAGTGCAGGTCGTGATCTCTATAAGACGTTATT	10500
K M L G A T V V S L A R S D E F V P I D CAAAATGTTGGGTGCTACAGTTAGTTAGCAAGGAGCGACGAATTGTTCTATTGA	10560
T E A V S E D D R N K A I T W A K K Y Q TACTGAAGCTGTAAGTGAAGATGATAGAAATAAGCAATCACATGGCAAAAAAATATCA	10620
L D A I F S T D G D G D R P L I A D E Y GTTAGATGCTATATTTCACACTGATGGTGTGGAGATCGCCCTCTGATAGCTGACGAATA	10680
G N W L R G D I L G L L C S L E L A A D TGGAAATTGGTTAAGAGGAGATATTAGGCCTCTGCTCTCGAATTAGCTGCTGA	10740
A V A I P V S C N S T I S S G N F F K H TGCAGTCGCTATTCCGTAAAGCTGCAACAGTACAATCTCATCTGGTAACTTTTAAACA	10800
V E R T K I G S P Y V I A A F A K L S A TGTGGAACGAACAAAGATTGGTCACCCCTATGTGATTGCAGCATTTGCTAAATTATCTGC	10860
N Y N C I A G F E A N G G F L L G S D V AAACTATAATTGTATAGCTGGTTTGAAACGAATGGTGGCTCTGCTAGGTAGCGATGT	10920
Y I N Q R L L K A L P T R D A L L P A I TTATATTAAATCAGCGTTACTTAAGGCATTACCAACACGTGATGCTTATTACCTGCCAT	10980
M L L F G S K D K S I S E L V K K L P A TATGCTCTGTTGGTAGCAAGGACAAAGTATTAGTGAGCTGTTAAAAACTTCCTGC	11040
R Y T Y S N R L Q D I S V K T S M S L I TCGCTATAACCTATTCAAAACAGATTACAGGATATAAGTGTAAAACAAGTATGTCTTAAT	11100
N L G L T D Q E D F L Q Y I G F N K H H AAATCTGGTCTGACAGATCAAGAGGATTTTGCACTATATTGGTTAATAAACATCA	11160
I L H S D V T D G F R I T I D N N N I I TATATTACATTCTGATGTTACTGATGGCTTGAATCACTATCGATAACAAACAATTAT	11220
H L R P S G N A P E L R C Y A E A D S Q TCATTTACGACCTTCAGGCAATGCCCTGAGTTGCCTGCTATCGGGAGGCTGACTCGCA	11280
E D A C N I V E T V L S N I K S K L G R AGAGGATGCATGTAATATTGTTGAAACTGTTCTCTCAATATCAAAGCAAACGGTAG	11340
End of manB	
A *	
AGCTTAATGCTGTTGATAATAGAGCGTTCTTCAGTAATACTTGTCTGGTTATCTGG	11400
Start of wbaP	
M D R F D N K Y N P N L TACCCAAGTTGAGGGTGAGAATTAA <u>ATGGATCGTTTGATAATAAGTATAACCCAAATT</u>	11460
C K I L L A I S D L L F F N V A L W A S ATGCAAAATATTATGGCTATATCAGATTTACTGTTTTAATGTAGCCTATGGCCTC	11520

Figure 9/10

L G V V Y L I F D E V Q R F V P Q E Q L		
GTAGGAGTTGTATATTAATCTTGATGAAGTCAGCGATTGTACCACAAGAGCAATT	11580	
D N R F I S H F I L S I V C V G W F W V		
AGATAATCGATTTATACATCTATCTATAGTATGCCTGGATGGTTGGGT	11640	
R L R H Y T Y R K P F W Y E L K E V I R		
TCGACTGCGTCACTATACATATCGAAAGCATTCTGGTATGAGTTGAAAGAGGTTATCG	11700	
T I V I F A V F D L A L I A F T K W Q F		
TACTATCGTTATTTGCTGTGTTGATTGGCTTAATTGCCTTACAAATGGCAGTT	11760	
S R Y V W V F C W T F A I I L V P F F R		
TTCACGCTATGCTGGGTGTTTGTTGACTTTGCCATAATCCTGGTGCCTTTTCG	11820	
A L T K H L L N K L G I W K K K T I I L		
CGCACTTACAAAGCATTATTGAAACAAGCTAGGTATCTGGAAAGAAAAAAACTATCCT	11880	
G S G Q N A R G A Y S A L Q S E E M M G		
TGGGAGCGGACAGAATGCTCGTGGTGCATATTCTGCCTGCAAAGTGGAGATGATGGG	11940	
F D V I A F F D T D A S D A E I N M L P		
GTGGATGTTATCGCTTTTTGATACGGATGCGTCAGATGCTGAAATAATATGTTGCC	12000	
V I K D T E T I W D L N R T G D V H Y I		
GGTGATAAAGGACACTGAGACTATTTGGGATTAAATCGTACAGGTATGTCATTATAT	12060	
L A Y E Y T E L E K T H F W L R E L S K		
CCTTGCTTATGAATACACCGAGTTGGAGAAAACACATTGGCTACGTGAACCTTCAA	12120	
H H C R S V T V V P S F R G L P L Y N T		
ACATCATTGTCGTTCTGTTACTGTCGTCCTCGTTAGAGGATTGCCATTATAAAC	12180	
D M S F I F S H E V M L L R I Q N N L A		
TGATATGTTCTTATCTTAGCCATGAAGTTATGTTATTAAGGATACAAAATACTGGC	12240	
K R S S R F L K R T F D I V C S I M I L		
TAAAAGGTCGTCCTGGTCTCAAACGGACATTGATATTGTTGTTCAATAATGATTCT	12300	
I I A S P L M I Y L W Y K V T R D G G P		
TATAATTGCATACCAACTTATGATTATCTGTTGAGTAAAGTTACTCGAGATGGTGGC	12360	
A I Y G H Q R V G R H G K L F P C Y K F		
GGCTATTATGGTCACCAAGCGAGTAGGTCGGCATGAAAACCTTTCCATGCTACAAATT	12420	
R S M V M N S		
TCGTTCTATGGTTATGAATTC	12441	

GAATTGGGAGGCGCAATGAAAGTCAGCTTTCTGCTGAAATTCCACTCTCATCGGA	60
AACCTTGTGCTGAATCAGATTACTGCAGTTATTGATATGGGCATGAGGTGGAGATTGT	120
CGCGTTACAAAAGGCGATACCCAACATACTCACGCCGCTGGGAGAAGTATGGCCTGGC	180
GGCGAAAACCCGCTGGTTACAGGATGAGCCCCAGGGACGGCTGGCGAAACTGCGCTACCG	240
GGCATGTAAAACGCTGCCGGGCTGCATCGGCGGCGACCTGGAAAGCGCTCAATTAC	300
CCGCTATGGCGATGAATCACGCAATTGATCCTTCCCGATTTGCGCGCAGGTGAGCCA	360
GCCTTTGTGGCGATGTGTTATCGCACACTTGGTCCGGCGGGCGTGACGGCGGCCAA	420
ACTACGCGAACTGGCGTGCTTCGCGAAAATCGCGACTATTTCACGGGATTGATAT	480
CTCTAGTCGTGAGGTGCTCAGTCATTACACGCCGGAGTATCAGCAGTTGTTCGTCGTGG	540
CGATCTGATGCTGCCCATCAGCGATCTGTGGCCGGTCGCGCTGAAAAGTATGGCGTGTCC	600
GCCGGAAAAGATTGCCGTTCGCGCATGGCGTCGACATGACGCGTTTACCCATCGTTC	660
GGTGAAGCGCCAGGGATGCCGCTGGAGATGATTCCGTGCGCGCCTGACAGAAAAAA	720
AGGCCTGCATGTGGCGATTGAAGCCTGTCGGCACTGAAAGCACAGGGCGTGGCGTTCG	780
CTACCGCATTCTGGGATTGGCCCGTGGGAACGTGGCTGCGCACGCTCATCGAGCAGTA	840
TCAGCTAGAGGATGTCATTGAGATGCCGGGTTAAACCGAGCCATGAAGTGAAGGCAT	900
GCTGGATGACGCCGATGTTTTGCTGCCGTGCGATTACCGGTACGGATGGCGATATGGA	960
AGGTATTCCGGTAGCGCTGATGGAGGCGATGGCGGTAGGGATTCCGTGTATCTACCGT	1020
GCATAGCGGTATTCCGGAACCTGGTGGAGGCCGGAAATCCGGCTGGCTGGTGCGGGAAA	1080
CGATGCCAGGCGCTGGCGCCCGACTCGCTGAGTCAGCCGGATTGACCACGACACGCT	1140
GGAGTCGGTGATCACGCGCCCGTGAAAAAGTGGCGCAAGATTAAATCAGCAGGGCAT	1200
TAATGCCAGTTAGCCAGCCTGCTACAAACGATATAACGAGGTGGTATGCCCGCACTA	1260
AATTCTCCGACGTACCCCTCTGACGGCAGGTTCTGCGCTTGCTGTTCTCCTTTCTGC	1320
GCGCCTTGCCTGGTACAGGCGCGTGAAACCTCGCGAGACCGTCGATATTAAGGATTATCCGG	1380
CGGATGACGGTATGCCCTCGTTCAAACAGGCCCTCGCCGACGGACAGACCGTGGTCGTAC	1440
CGCCAGGATGGGTGTGTAAAATATCAATGCGCGATAACGATTCCGGCGGGAAAACGC	1500
TGCGGGTACAGGGCGCGTGGGAATGGCGGGGACGGTTTATTTCAGGACGGGT	1560
GTCAGGTGGTGGGGGAGCAGGGCGGCAGTCTGCACAATGTGACGCTGGATGTTCGCGGGT	1620
CGGACTGTGTGATTAAGGCGTGGCGATGAGCGGTTGGCCCCGTCGCGCAAATTTC	1680
TCGGTGGTAAGGAACCGCAGGTGATGCCATTCTCAGGATTTCAAAACGATGGCGCGCGGATTA	1740
ACGCCAACTACGCCATTCTCCGCCAGGATTTCAAAACGATGGCGCGCGGATTA	1800
CGCATAGCCGTTAGCGATTACAGGGGACGCCATTGAGTGGAAATGCGCGATTACG	1860
ACCGCGACATCCTGATTCCGATCATGTCATCGAACGCATTAATTGTACCAATGGCAAA	1920
TCAACTGGGGATCGGCATCGGCTGGCGGGTAGCACCTATGACAACAGTTATCCTGAAG	1980

Figure 10/1

ACCAGGCAGTAAAAACTTGTGGTGGCAATATTACCGGATCTGATTGCCGACAGCTTG	2040
TGCACGTAGAAAATGGCAAACATTCGTCAATCGCAATGTCAAAGCAAAAACATCACGC	2100
CCGGTTTCACTAAAAATGCCGGTATTGATAACGCAACGATCGCAATTATGGCTGTGATA	2160
ATTCGTCAATTGATAATATTGATATGACGAATAGTGCCGGATGCTCATCGGCTATGGCG	2220
TCGTTAAAGGAAAATACCTGTCAATTCCGCAAAACTTTAAATTAAACGCTATTGGTTGG	2280
ATAATGCCAGGTTGCTTATAAATTACGCCATTCAAATTCCCTCCGGAACACCCCCCT	2340
CTTTTGTGCCATCACCAATGTACGGATGACGCGTGCTACGCTGGAACTGCATAATCAAC	2400
CGCAGCACCTCTTCTGCGCAATATCACCGTGTGATGCAAACCTCAGCGATTGGCCCGCGT	2460
TAAAAATGCATTTGCGATTGCGTAAAGATGTACGTGGTCAATTATGGCCGCCAGGACA	2520
CGCTGCTTCCCTCGCTAATGTCATGCCATCAATGAAAACGGGCAGAGTTCCGTGGATA	2580
TCGACAGGATTAATCACCAAACCGTGAATGTCGAAGCAGTGAATTTCGCTGCCGAAGC	2640
GGGGAGGGTAAGTACCGCTATTTCAGAAAATTCTGGAAAAAGTTGTTCATACTTAA	2700
TGTTATGGTGGCGACTAAGACGTAATGTAGAGCGTGCCATCATTATCCCTGGCAGCAGAG	2760
TAATTCATGCTGGCGAAAACAAGCTAAAGAGCTATAATTCAAGCAACCATTACAGGTGG	2820
AAGAAACAATGATGAATTGAAAGCAGTTACCGGTAGCGGTTGGGTATGCATATGT	2880
TGCCTGCCACCAAGGCAATCCAAAAGAGATGCTACCGATCGTCGACAAGCCAATGATTC	2940
AGTACATTGTCGATGAGATTGTCAGGATCAAAGAAATCGTGTGGTACTCACG	3000
CGTCTAAAACGCCGTGAGAACCACTTCGACACCTCTTATGAACCTGAATCACTCTTG	3060
AGCAGCGCGTTAACCGTCAGCTTTGGCGGAAGTGCATCTATCTGCCAACGGCGTGA	3120
CGATTATGAACGTTGCCAGGCGCAGCGTTAGGGCTGGGCATTCTATTCTGTGCCGC	3180
GTCCGGTGTGGCGATAACCTTCAATGTTGACTCCGGATATTATTATCGATGATG	3240
CTACCGCCATCCGCTGCGCTATAACCTGCGCATGGTGGCGCTTCAATGAAACGG	3300
GTCGCAGCCAGGTGCTGGCGAAGCGCATGAAAGGTGATTATCGGAGTATTCCGTTATCC	3360
AGACGAAAGAACCTCTGGATAATGAAGGCAAAGTCAGCCGGATTGTGGAGTTATCGAAA	3420
AACCGGATCAGCCGAGACGCTGGATTCCGATTTGATGGCGGTAGGCCGTTATGTGCTTT	3480
CAGCCGACATCTGGCGGAACCTGGAAAGAACCGAACCGGGCCCTGGGCCGATCCAGC	3540
TCACCGATGCCATTGCTGAACCTGGCGAAAAACAGTCGGTTGACGCGATGCTAATGACGG	3600
GTGACAGCTATGACTGCGTAAAAAAATGGCTACATGCAGGCATTGTGAAGTACGGC	3660
TGCGCAACCTGAAAGAAGGAGCCAAGTCCGTAAGAGCATAGAGCAGCTTTGATGAAT	3720
AAGTATTAACAACCGTGATAATGGTTGGTGTAAACATAATAACGGCAGTGAACATTG	3780
AAGCGGCAAGTTGGCTGAAACGAGTGTGACTGCCGTTAGTTGTATAAAGGGCTTA	3840
AGTAACAAGGGTTATCTGGAGCATTTAATGCTGATTTATAAGATTAATCCTTGTTC	3900
CGGATGCAATTAAAGACAATTAGCGTTAAGTTAGTGAAGCTTGCCCTGCTGGCG	3960

Figure 10/2

AGGTTTGCACAAAGTCGATATGTACGCAGTGCAGTGGTAGCTGATGAGCCAGGGCGGTA 4020
 GCGTGTGTAACGACTTGAGCAATTAAATTATTGGCAAATTAAATACCACATTAAATAC 4080

Start of *rmlB*
 V K I L I T G G A G F I G S
 GCCTTATGGAATAGAAAAGTGAAGATACTTATTACTGGCGGGCAGGTTTATTGGATCA 4140

 A V V R H I I K N T Q D T V V N I D K L
 GCTGTTGTCGCCATATTATTAAGAACACAGGACACTGTAGTTAATATTGATAAAATTA 4200

 T Y A G N L E S L S D I S E S N R Y N F
 ACCTACGCCGTAATCTGAATCCCTTCTGATATTCTGAAAGTAATCGCTACAATTT 4260

 E H A D I C D S A E I T R I F E Q Y Q P
 GAACACGCCGATATTGTGATTCGCTGAAATAACCGTATTGAGCAGTACCGAGCCG 4320

 D A V M H L A A E S H V D R S I T G P A
 GACGCCGTGATGCATTGGCTGCCAAAGTCATGTGGACCGTCATTACCGGGCCAGCA 4380

 A F I E T N I V G T Y A L L E V A R K Y
 GCATTTATTGAAACCAATATCGCGCACCTATGCACTTCTGAAAGTTGCGCGTAAATAC 4440

 W S A L G E D K K N N F R F H H I S T D
 TGGTCTGCCCTGGCGAAGATAAAAAAAATAATTCGTTTCATCATATTCACGTGAT 4500

 E V Y G D L P H P D E V E N S V T L P L
 GAAAGTTACGGCGATTACCGCATCCTGATGAAGTTGAAACAGCGTTACGCTGCCGTTA 4560

 F T E T T A Y A P S S P Y S A S K A S S
 TTTACTGAAACGACGCCATATGCCCAAGTAGCCCCTATTCTGCGTCAAAAGCATCCAGC 4620

 D H L V R A W R R T Y G L P T I V T N C
 GATCATTAGTCGCTGCCCTGGCGCGTACCTATGGTCTACCAACGATCGTTACCAATTGT 4680

 S N N Y G P Y H F P E K L I P L V I L N
 TCTAATAACTATGCCCTTATCACTCCCTGAAAAACTGATTCCGTTGGTCATTTGAAC 4740

 A L E G K P L P I Y G K G D Q I R D W L
 GCACCTGGAAGGAAAGCCTTGCCAATTATGGCAAAGGGATCAGATTCCGATTGGCTA 4800

 Y V E D H A R A L H M V V T E G K A G E
 TATGTAGAAGATCATGCTCGCGCTTCATATGGTAGTGACTGAAGGCAAGGCAGGGAG 4860

 T Y N I G G H N E K K N L D V V F T I C
 ACTTATAACATTGGTGGACACAATGAGAAGAAAACTCGATGTGGTATTTACCATCTGT 4920

 D L L D E I V P K A T S Y R E Q I T Y V
 GATCTGCTGGATGAGATTGTACCCAAAGCGACTTCTATCGTAACAAATCACTTATGTC 4980

 A D R P G H D R R T Y A I D A G K I S R E
 GCGGATCGTCCGGGCCATGATCGCTTATGCCATTGATGCAGGTAAAATTAGCCCGAA 5040

 L G W K P L E T F E S G I R K T V E W Y
 TTAGGCTGGAAACCGCTGGAGACCTTGAAAGCGGTATTGTAACAGTGGAAATGGTAC 5100

 L A N T Q W V N N V K S G A Y Q S W I E
 CTTGCAAATACTCAATGGTAAACAAATGTTAAAAGTGGCGTATCAGAGTTGGATAGAA 5160

End of *rmlB* Start of *rmlD*
 Q N Y E G R Q * M N I L L F G K T G Q V
 CAGAACTATGAAGGACGCCAG TAATGAATATCTTACTTTGGTAAGACAGGGCAAGTAG 5220

Figure 10/3

G W E L Q R S L A P V G N L I A L D V H GCTGGGAGTTGCAACGTTCTCTGGCACCGGTAGGAATCTGATTGCCCTGGATGTCCATT	5280
S K E F C G D F S N P K G V A E T V R K CAAAAGAGTTTGCAGGTAGTTAGTAATCCGAAAGGCAGGTGCCGAAACCGTTCGTAAGC	5340
L R P D V I V N A A A H T A V D K A E S TTCGTCCCGATGTGATTGTTAACGCAGCAGCCCATACTGCAGTAGATAAAAGCAGAGTCTG	5400
E P E L A Q L L N A T S V E A I A K A A AACAGAACTGGCGCAGTTACTAACGCCACCAAGTGTGGAAGGCATCGCTAAAGCAGCCA	5460
N E T G A W V V H Y S T D Y V F P G T G ACGAAACTGGCGCATGGTAGTGCATTATCAACCGATTATGTATTCCTGGTACCGGCG	5520
D I P W Q E T D A T S P L N V Y G K T K ATATCCCATGGCAGGAAACGGACGCTACGTCGCCGTGAATGTCTATGGCAAAACCAAC	5580
L A G E K A L Q D N C P K H L I F R T S TGGCGGGAGAAAAGGCCCTGCAGGATAACTGCCCTAAACACCTATCTTCCGCACCAGTT	5640
W V Y A G K G N N F A K T M L R L A K E GGGTTTATGCAGGTAAGGGCAATAATTGCAAAGACAATGCTTCGTCTGGCGAAAGAGC	5700
R Q T L S V I N D Q Y G A P T G A E L L GTCAGACACTTCAGTCATTAACGATCAGTACGGTGCACCCGGTGCAGGAAATTACTGG	5760
A D C T A H A I R V A L N K P E V A G L CTGACTGTACGGCGCATGCGATCCGTGTGGCGTTAAATAACAGAAGTCGCAGGTCTTT	5820
Y H L V A G G T T T W H D Y A A L V F D ACCATCTGGTTGCCGGGGAACCAACCTGGCATGACTACGGCCCTTAGTCTTGACG	5880
E A R K A G I T L A L T E L N A V P T S AGGCGCGCAAAGCAGGGATAACGCTTGCCTGACTGAGCTTAATGCTGTGCCGACCAGCG	5940
A Y P T P A S R P G N S R L N T E K F Q CCTACCCGACGCCGGCAGCAGACCAGGCAATTGCGTCTCAATACTGAAAAGTTTCAGC	6000
R N F D L I L P Q W E L G V K R M L T E GTAATTTGACCTTATTCTGCCCTCAATGGGAATTAGGAGTTAACCGTATGCTGACTGAAA	6060
End of rmlD	
M F T T T T I * TGTGTTACGACGACAACCATCTAATAAATTAAATGCCCATCAGGGCATTTCATGAATG	6120
Start of rmlA	
M K T R K G I I L A G G S G T R L AGAAATGGAAATGAAAACGCAGGTAGGGCATTATTAGCGGGGGCTCCGGCACCCGTCT	6180
Y P V T M A V S K Q L L P I Y D K P M I TTATCCGGTGACCATGGCGTAAGTAAGCAATTGCTACCAATTATGATAAAACCGATGAT	6240
Y Y P L S T L M L A G I R D I L I I S T TTACTATCCCTTCCACGCTATGCTGCAGGCATTGGGATATCCTGATCATCAGTAC	6300
P Q D T P R F Q Q L L G D G S Q W G L N GCCACAGGACACGCCCGTTCAACAACTGCTGGGAGACGGCAGCCAGTGGGGCTGAA	6360
L Q Y K V Q P S P D G L A Q A F I I G E TCTTCATATAAGTACAGCCAAGCCGGATGGCTAGCACAGGCAGTATTATTGGTGA	6420
E F I G H D D C A L V L G D N I F Y G H AGAGTTCATGGTCATGATGATTGTGCATTAGTGCCTGGTACAATATCTTCTATGGTCA	6480

D L P K L M E A A V N K E S G A T V F A 6540
 TGATTTACCAAAGTTAATGGAAGCTGCCGTTAATAAAGAAAGTGGTGTACCGCTTCGC

 Y H V N D P E R Y G V V E F D Q K G T A 6600
 TTATCATGTAAACGA'TCCGGAGCGCTACGGTGTGGTTGAGTTGACCAAAAGGGCACAGC

 V S L E E K P L Q P K S N Y A V T G L Y 6660
 CGTTAGTCTGGAAGAAAAACCATTACAACCGAAGAGTAATTACCGGTAACGGGCTGTA

 F Y D N S V V E M A K N L K P S A R G E 6720
 TTTTATGATAATAGCGTGGTGGAGATGGCGAAAATCTTAAGCCTCCGCTCGCGGTGA

 L E I T D I N R I Y M E Q G R L S V A M 6780
 GTTAGAAATCACGGATATTAACCGTATCTATATGGAGCAGGGAAAGATTGTCTGTCGCTAT

 M G R G Y A W L D T G T H Q S L I E A S 6840
 GATGGGGCGCGGTTATGCCTGGCTGGATACAGGGACGCATCAGAGTTGATAGAGGCCAG

 N F I A T I E E R Q G L K V S C P E E I 6900
 TAATTTATTGCAACCATCGAAGAACGCCAGGGCTAAAAGTGCCTGCCCGGAAGAGAT

 A F R K N F I N A Q Q V I E L A G P L S 6960
 CGCATTTCTGTAAGGTTATAGCACAACAGTTATAGAACTGGCCGGGCCATTATC

End of rmlA Start of rmlC
 K N D Y G K Y L L K M V K G L * V M I V 7020
 AAAAATGATTATGCCAAATATTGCTGAAGATGGTGAAGGTTA TAAGTGATGATTGT

 I K T A I P D V L I L E P K V F G D E R 7080
 GATTTAAACAGCAATACCAGATGTCTTGATCTTAGAGCCTAAAGTTTGGCGATGAGAG

 G F F F E S Y N Q Q T F E E L I G R K V 7140
 GGGATTCTTTGAAAGTTATAACCAGCAGACCTTGAGAGTTGATTGGACGTAAAGT

 T F V Q D N H S K S K K N V L R G L H F 7200
 TACATTTGTTCAAGATAATCATTCAAAATCCAAAAGAACGTACTCAGAGGGCTACATT

 Q R G E N A Q G K L V R C A V G E V F D 7260
 TCAGAGAGGAGAAATGCACAGGGGAAGTTAGTCGTTGTGCTGCGTGGTTTTGA

 V A V D I R K E S P T F G Q W V G V N L 7320
 TGTTGCGGTCGATATCCGAAAAGAACCGCTACTTTGGTCAATGGGTTGGTAAATCT

 S A E N K R Q L W I P E G F A H G F V T 7380
 GTCTGCTGAGAATAAGCGACAGCTTGATTCCAGAAGGTTTGCTCATGGTTTGTAC

 L S E Y A E F L Y K A T N Y Y S P S S E 7440
 TCTTAGTGAGTATGCAGAGTTCTGTACAAAGCAACTAATTATTACTCACCTTCATCGGA

 G S I L W N D E A I G I E W P F S Q L P 7500
 AGGTAGCATTCTATGGAATGATGAGGCAATAGGTATTGAATGGCCTTTCTCAGCTGCC

End of rmlC
 E L S A K D A A A P L L D Q A L L T E * 7560
 TGAGCTTCAGCAAAGATGCTGCAGCACCTTACTGGATCAAGCCTGTTAACAGAGTA

Start of ddhD
 V S H I I K I F P S N I E F S G R E 7620
 AGCATCGTGTCTCATATTATTAAGATTTCCATCAAATATTGAATTTCCGGTAGAGAG

 D E S I L D A A L S A G I H L E H S C K 7680
 GATGAATCAATCCTCGATGCTGCCTATCGGCTGGTATCCATCTGAACATAGCTGCCAAA

 A G D C G I C E S D L L A G E V V D S K 7740
 CGGGGTGATTGTGGTATCTGTGAGTCCGATTGTTGGCGGGAGAAGTTGTTGACTCCAAA

Figure 10/5

G N I F G Q G D K I L T C C C K P K T A		
GGTAATATTTGGACAGGGTGATAAAATACTAACCTGCTGCTGTAACCTAAACGCC	7800	
L E L N A H F F P E L A G Q T K K I V P		
CTTGAGCTAAATGCGCATTTCCTGAACTAGCTGGACAGACAAAAAAATTGTCCCA	7860	
C K V N S A V L V S G D V M T L K L R T		
TGCAAGGTAAATAGTGCTGTACTGGTTCAAGGCATGTTATGACTTGAAGTTACGCACA	7920	
P P T A K I G F L P G Q Y I N L H Y K G		
CCACCAACAGCAAAATTGGCTCCTCCAGGGCAGTATATCAATTACATTATAAGGT	7980	
V T R S Y S I A N S D E S N G I E L H V		
GTAACTCGCAGTTATTCTATCGCTAATAGTGTAGTCGAATGGTATTGAGTGCATGTA	8040	
R N V P N G Q M S S L I F G E L Q E N T		
AGGAATGTTCCAATGGTCAGATGAGTCGCTCATTGGGGAGTTACAAGAAAATACT	8100	
L M R I E G P C G T F F I R E S D R P I		
CTTATGCGCATTGAAGGGCCTTGCAGAACATTTCATTGAAAGTGACAGACCTATA	8160	
I F L A G G T G F A P V K S M V E H L I		
ATCTTCCTGCAGGCGGTACTGGATTGCTCCAGTTAAATCAATGGTTGAGCATCTCATT	8220	
Q G K C R R E I Y I Y W G M Q Y S K D F		
CAGGGAAAATGTCGTCGTGAGATCTACATTACTGGGAATGCAATATAGTAAAGATTT	8280	
Y S A L P Q Q W S E Q H D N V H Y I P V		
TACTCTGCATTACCGCAGCAGTGGAGTGAACAGCACGACAACGTTCATTATATCCCTGTT	8340	
V S G D D A E W G G R K G F V H H A V M		
GTTTCTGGTGATGACGCCAATGGGGGGAAAGAAAGGGATTGTCATCATGCCGTGATG	8400	
D D F D S L E F F D I Y A C G S P V M I		
GATGATTGATTCTAGAGTCTCGATATATGATGTGGTCACCTGTGATGATC	8460	
D A S K K D F M M K N L S V E H F Y S D		
GATGCCAGTAAAAGGACTTATGATAAAATCTCTGTAGAACATTCTATTCTGAT	8520	
End of ddhD Start of ddhA		
A F T A S N N I E D N L *		
M K A V I L A G		
GCATTTACCGCATCTAATAATATTGAGGATAATTATGAAAGCGGTATCCTGGCTGGTG	8580	
G L G T R L S E E T I V K P K P M V E I		
GACTTGGTACCACTAAGTGAAGAAACAATTGTAAAACCAACCGATGGTAGAAATTG	8640	
G G K P I L W H I M K M Y S V H G I K D		
GTGGCAAGCCTATTCTTGGCACATTATGAAAATGTATTCTGTGATGGTATCAAGGATT	8700	
F I I C C G Y K G Y V I K E Y F A N Y F		
TTATTATCTGCTGTGGTATAAAGGATATGTGATTAAGAAATTGGCAACTACTTCC	8760	
L H M S D V T F H M A E N R M E V H H K		
TTCACATGTCAGATGTAACATTCCATATGGCTGAAAACCGTATGGAAGTTCACCATAAAC	8820	
R V E P W N V T L V D T G D S S M T G G		
GTGTTGAACCATGGAATGTCACATTGGTGATACGGGTGATTCTCAATGACTGGTGGTC	8880	
R L K R V A E Y V K D D E A F L F T Y G		
GTCTGAAACGTGTTGCTGAATACGTAAAAGATGACGAGGCTTCCTGTTACTTATGGTG	8940	
D G V A D L D I K A T I D F H K A H G K		
ATGGCGTTGCCGACCTTGATATCAAAGCGACTATCGATTCCATAAGGCTCACGGTAAGA	9000	

Figure 10/6

K A T L T A T F P P G R F G A L D I R A AAGCGACTTTAACAGCTACTTTCCACCAGGACGCTTGGCGCATTAGATATCCGAGCTG	9060
G Q V R S F Q E K P K G D G A M I N G G GTCAGGTCCGGTCATTCAGGAAAAACCGAAAGGCATGGGCAATGATCAATGGTGGTT	9120
F F V L N P S V I D L I D N D A T T W E TCCTTGTTGTTGAATCCATCGGTTATCGATCTCATCGATAACGATGCAACAACCTGGAAC	9180
Q E P L M T L A Q Q G E L M A F E H P G AAGAGCCATTAATGACATTGGCACAAACAGGGGGAGTTAATGGCTTTGAACACCCAGGTT	9240
F W Q P M D T L R D K V Y L E G L W E K TCTGGCAGCCGATGGATACCTACGTGATAAAGTTACCTCGAAGGGCTGTGGGAAAAAG	9300
End of <i>ddhA</i> Start of <i>ddhB</i>	
M I D K N F W Q G	
G K A P W K T W E * GTAAAGCTCCGTGGAAAACCTGGAG TAACTAGATGATTGATAAAAATTTGGCAAGGT	9360
K R V F V T G H T G F K G S W L S L W L AAACGTGTATTGTTACCGGCCATACTGGCTTAAAGGAAGCTGGCTTCGCTATGGCTG	9420
T E M G A I V K G Y A L D A P T V P S L ACTGAAATGGGTGCAATTGTAAAAGGCTATGCACTTGATGCGCCAATGTTCCAAGTTA	9480
F E I V R L N D L M E S H I G D I R D F TTTGAGATAGTGCCTTAATGATCTATGGAATCTCATATTGGCGACATTGCGATTGTT	9540
E K L R N S I A E F K P E I V F H M A A GAAAAGCTGCGCAATTCTATTGAGAATTAAAGCCAGAAATTGTTCCATATGGCAGCC	9600
Q P L V R L S Y E Q P I E T Y S T N V M CAGCCTTAGTGCGCCTATCTTATGAACAGCCAATCGAAACATACTCAACAAATGTTATG	9660
G T V H L L E T V K Q V G N I K A V V N GGTACTGTCCATTGCTTGAAACAGTTAACGCAACTAGGTAACATAAAGGCAGTCGTAAAT	9720
I T S D K C Y D N R E W V W G Y R E N E ATCACCAAGTGATAAGTGCCTACGACAATCGTAGTGGGTGTGGGCTATCGTGAGAACGAA	9780
P M G G Y D P Y S N S K G C A E L V A S CCCATGGGAGGGTACGATCCATACTCTAACAGTTAACAGTTAGTCGCGTCT	9840
A F R N S F F N P A N Y E Q H G V G L A GCATTCCGGAACTCATTCTCAATCCTGCAAATTATGAGCAACATGGCGTTGGCG	9900
S V R A G N V I G G G D W A K D R L I P TCTGTGAGGGCTGGTAATGTCATAGGCGGAGGCGATTGGCTAAAGACCGTTAACCT	9960
D I L R S F E N N Q Q V I I I R N P Y S I GATATTCTGCGCTATTGAAAATAACCAGCAGGTTATTATCGAAACCCATATTCTATC	10020
R P W Q H V L E P L S G Y I V V A Q R L CGTCCCTGGCAGCATGTAATGGAGCCTTTCTGGTTACATTGTTGGCGAACGCTTA	10080
Y T E G A K F S E G W N F G P R D E D A TATACAGAAGGTGCTAAGTTCTGAAGGATGGAATTGGCCCGCGTGAAGATGCG	10140
K T V E F I V D K M V T L W G D D A S W AAGACGGTCGAATTATTGTTGACAAGATGGTCACGCTTGGGTGATGATGCAAGCTGG	10200
L L D G E N H P H E A H Y L K L D C S K TTACTGGATGGTGAGAATCATCCTCATGAGGCACATTACCTGAAACTGGATTGCTCTAAA	10260

A N M Q L G W H P R W G L T E T L G R I 10320
 GCAAATATGCAATTAGGATGGCATCCGCGTTGGGGATTGACTGAAACACTTGGTCGCATC

 V K W H K A W I R G E D M L I C S K R E 10380
 GTAAAATGGCATAAAGCATGGATTGCGCGAAGATATGTTGATTGTTCAAAGCGTGAA

End of ddhB
 I S D Y M S A T T R * 10440
 ATCAGCGACTATATGTCCTGCAACTACTCGTTAAGAAAATAAGTTAAGGAATCAAAGTAA

Start of ddhC
 M T A N N L R E Q I S Q L V A Q Y A N E 10500
TGACAGCAAATAACCTGCGTGAGCAAATCTCAGCTTGTCGCTCAGTATGCGAATGAGG

 A L S P K P F V A G T S V V P P S G K V 10560
 CATTGAGCCGAAACCTTTGTTGCAGGTACAAGCGTTGTCCTCCGGAAAGGTTA

 I G A K E L Q L M V E A S L D G W L T T 10620
 TTGGTGCCAAAGAGTTACAATTGATGGTTGAGGCGTCTCTGATGGATGGCTAACTACTG

 G R F N D A F E K K L G E F I G V P H V 10680
 GTCGTTCAATGATGCCTTGAAAAAAACTTGGGAATTATTGGGTTCTCATGTTT

 L T T T S G S S A N L L A L T A L T S P 10740
 TAACGACAACATCTGGCTCTCGGCAAATTGCTGGCACTGACTGCGCTGACTTCCCCAA

 K L G E R A L K P G D E V I T V A A G F 10800
 AATTAGGCAGCGAGCTCTCAAACCTGGTGATGAGGTTATTACTGTCGCTGCTGGCTTCC

 P T T V N P A I Q N G L I P V F V D V D 10860
 CGACTACAGTTAACCGGGCATCCAGAATGGTTAACCGGTATTGCGATGTTGATA

 I P T Y N I D A S L I E A A V T E K S K 10920
 TCCCGACATATAATATCGATGCCTCTCTCATTGAAGCTGAGTTACTGAGAAAATCAAAAG

 A I M I A H T L G N A F N L S E V R R I 10980
 CGATAATGATCGCTCATACACTCGGTAATGCATTAAACCTGAGTGAAGTTCGTGGATTG

 A D K Y N L W L I E D C C D A L G T T Y 11040
 CCGATAAAATATAACTTATGGTGATTGAAGACTGCTGTGATGCCCTGGGACGACTTATG

 E G Q M V G T F G D I G T V S F Y P A H 11100
 AAGGCCAGATGGTAGGTACCTTGGTGACATCGGAACCGTTAGTTTATCCGGCTCACC

 H I T M G E G G A V F T K S G E L K K I 11160
 ATATCACAATGGGTGAAGGGCGGTGCTGTATTCAACCAAGTCAGGTGAACTGAAGAAAATTA

 I E S F R D W G R D C Y C A P G C D N T 11220
 TTGAGTCGTTCCGTGACTGGGGCCGGATTGTTATTGTCGCCAGGATGCGATAACACCT

 C G K R F G Q Q L G S L P Q G Y D H K Y 11280
 CGGGTAAACGTTGGTCAGCAATTGGGATCACTCCTCAAGGCTATGATCACAAATATA

 T Y S H L G Y N L K I T D M Q A A C G L 11340
 CTTATCCCACCTCGGATATAATCTCAAAATCACGGACATGCAGGCAGCATGTGGCTGG

 A Q L E R V E E F V E Q R K A N F S Y L 11400
 CTCAGTTGGAGCGCGTAGAAGAGTTGTTAGAGCAGCGTAAAGCTAACTTTCCCTATCTGA

 K Q G L Q S C T E F L E L P E A T E K S 11460
 AACAGGGCTTGCAATCTGCACTGAATTCCCTCGAATTACCAAGAAGCAACAGAGAAATCAG

 D P S W F G F P I T L K E T S G V N R V 11520
 ATCCATCCTGGTTGGCTCCCTATCACCCCTGAAAGAAACTAGCGGTGTTAACCGTGTGCG

E L V K F L D E A K I G T R L L F A G N
 AACTGGTGAATTCTGATGAAGCAAAATCGGTACACGTTACTGTTGCTGGAAATC 11580

L I R Q P Y F A N V K Y R V V G E L T N
 TGATTGCAACCGTATTTGCTAATGTGAAATATCGTAGTGGGTGAGTTGACAAATA 11640

T D R I M N Q T F W I G I Y P G L T T E
 CCGACCGTATAATGAATCAAACGTTCTGGATTGGTATTTATCCAGGCTGACTACAGAGC 11700

End of ddhC

H L D Y V V S K F E E F F G L N F *
 ATTTAGATTATGTAGTTAGCAAGAGTTCTGGTTGAATTTC TAATTCAATT 11760

Start of abe

M T F L K E Y V I V S G A
 TATTCTATCTGGTGATGCGATGACCTTTGAAAGAATATGTAATTGTCAGTGGGCTT 11820

S G F I G K H L L E A L K K S G I S V V
 CCGGCTTATTGGTAAGCATTACTCGAAGCGCTAAAAAAATCGGGGATTCAGTTGTCG 11880

A I T R D V I K N N S N A L A N V R W C
 CAATCACTCGAGATGTAATAAAAATAATAGTAATGCATTAGCTAATGTTAGATGGTGCA 11940

S W D N I E L L V E E L S I D S A L I G
 GTTGGGATAATATCGAATTATTAGTCGAGGAGTTATCAATTGATTCTGCATTAATTGGTA 12000

I I H L A T E Y G H K T S S L I N I E D
 TCATTCACTGGCAACAGAATATGGCATAAAACATCATCTCTCATAAATATTGAAGATG 12060

A N V I K P L K L L D L A I K Y R A D I
 CAAATGTTATAAAACCATTAAAGCTTCTGATTGGCAATAAAATATCGGGCGGATATCT 12120

F L N T D S F F A K K D F N Y Q H M R P
 TTTTAAATACAGATAGTTTTGCCAAGAAAGATTTAATTATCAACATATGCGGCCCTT 12180

Y I I T K R H F D E I G H Y Y A N M H D
 ATATAATTACTAAAAGACACTTGATGAAATTGGCATTATTATGCTAATATGCATGACA 12240

I S F V N M R L E H V Y G P G D G E N K
 TTTCATTGTAAACATGCGATTAGAGCATGTATATGGCCTGGGATGGTAAATAAT 12300

F I P Y I I D C L N K K Q S C V K C T T
 TTATTCCATACATTATCGACTGCTTAAATAAAAACAGAGTTGCGTGAAATGTACAACAG 12360

G E Q I R D F I F V D D V V N A Y L T I
 GCGAACAGATAAGAGACTTATTTGATGAGTGTGGTAAATGCTTATTAACTATAT 12420

L E N R K E V P S Y T E Y Q V G T G A G
 TAGAAAATAGAAAAGAAGTACCTTCATATACTGAGTATCAAGTTGAACTGGTCTGGGG 12480

V S L K D F L V Y L Q N T M M P G S S S
 TAAGTTGAAAGATTTCTGGTTATTGCAAAATACTATGATGCCAGGTTCATCGAGTA 12540

I F E F G A I E Q R D N E I M F S V A N
 TATTGAAATTGGTGCAGAGCAAAGAGATAATGAAATAATGTTCTGTAGCAAATA 12600

N K N L K A M G W K P N F D Y K K G I E
 ATAAAATTTAAAAGCAATGGCCTGGAAACCAAATTGATTATAAAAAGGAATTGAAG 12660

End of abe

E L L K R L *
 AACTACTGAAACGGTTA TGAGATTTCATGATCTTAAATAAAATCGTTAACAAATT 12720

Start of wzx
 V K V Q L L
 AGTCGCGTTATGTTGAAAAACTAAGTCGTTAATTGCATAGTGAAAGTCATTGTTAA 12780

K I P S H L I V A G S S W L S K I I I A		
AAATTCCGAGTCATTTAATTGTTGCAGGTCATCATGGTATCCAAAATAATAATTGCCG	12840	
G V Q L A S I S Y L I S M L G E E K Y A		
GGGTGCAGTTAGCAAGTATTCATATCTTATTCTATGCTAGGTGAAGAGAAAATATGCAA	12900	
I F S L L T G L L V W C S A V D F G I G		
TCTTTAGTTGTTAACTGGTTATTAGTATGGTGTAGCGCTGTTGATTTGGCATAGGTA	12960	
T G L Q N Y I S E C R A K N K S Y D A Y		
CAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAACAAAAGTTATGATGCATATA	13020	
I K S A L H L S F I A I I F F I A L F Y		
TTAAATCAGCATTACATCTAACGCTTATAGCTATTATTTTTTATTGCTTATTTATA	13080	
I F S G V I S A K Y L S S F H E V L Q D		
TTTTTCTGGGTAATTCCGCTAAATATCTTCTTTCATGAGGTATTACAGGACA	13140	
K T R M L F F T S C L V F S S I G I G A		
AAACCAGAATGCTCTTTTACCTCATGTCGGTTTCAGTTCTATTGGAATCGGAGCTA	13200	
I A Y K I L F A E L V G W K A N L L N A		
TTGCTTATAAAACTTTTGCCTGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCAT	13260	
L S Y M I G M L G L L Y I Y Y R G I S V		
TATCTTATATGATAGGTATGCTCGCTTGCTATATATACTATAGGGGATCTCAGTTG	13320	
D I K L S L I V L Y L P V G M I S L C Y		
ACATAAAATTACTCAACTAACAGCTTATCATGTTAAAACAACAAAATCTCATTATAGCAA	13380	
I V Y R Y I K L Y H V K T T K S H Y I A		
TTGTATATAGATACTAAAGCTTATCATGTTAAAACAACAAAATCTCATTATAGCAA	13440	
I L R R S S G F F L F T L L S I V V L Q		
TTTACGTAGATCTCAGGGTTTTCTTTACTTATTATCGATAGTGGTCTTCAAA	13500	
T D Y M V I S Q R L T P A D I V Q Y T V		
CAGATTATGGTCATTCCTCAAAGGCTAACCTCGTGTGATATTGTCATACAGTAA	13560	
T M K I F G L V F F I Y T A I L Q A L W		
CGATGAAAATTGGTTAGTCTTTTACTGCTATTGCAAGCATTATGGC	13620	
P I C A E L R V K Q Q W K K L N K M I G		
CTATATGTGCTGAATTGAGAGTCACAAACAGCAATGGAAAAACTAACAAAATGATAGGTG	13680	
V N I L L G S L Y V V G C T I F I Y L F		
TCAATATTTGCTGGCTCACTATATGTTGGATGTACAATATTATTTATTTA	13740	
K E Q I F S V I A K D I N Y Q V S I L S		
AAGAACAGATATTCAGTAATAGCCAAAGATATTAATTCAAGTTCTATTATCTT	13800	
F M L I G I Y F C I R V W C D T Y A M L		
TTATGTTAATTGGCATATATTCTGTATTGCGTTGGTGTGACACTTATGCAATGTTAT	13860	
L Q S M N Y L K I L W I L V P L Q A I I		
TGCAAAGTATGAATTATTTAAACTTTGGATATTAGTACCAACTACAAGCAATAATTG	13920	
G G I A Q W Y F S S T L G I S G V L L G		
GTGGAATAGCACAATGGTATTTCTAGTACGCTGGAACTAGTGGAGTGCTGCTGGCT	13980	
L I I S F A L T V F W G L P L T Y L I K		
TGATTATATCTTTGCTTTAACTGTTTTGGGGCTTCCACTAACCTACTTAATTAGG	14040	

Figure 10/10

End of wzx Start of wbaV

A N K G * M L I - S F C I P T Y N R K Q	14100
CAAATAAGGGATAATCATATGCTTATATCATTGTATTCCAACCTATAATAGAAAACAA	
Y L E E L L N S I N N Q E K F N L D I E	14160
TATCTTGAAGAGTTGTTGAATAGTATAAATAATCAGGAAAAATTAAATTAGATATTGAG	
I C I S D N A S T D G T E E M I D V W R	14220
ATATGTATATCAGATAATGCCTACTGATGGTACAGAGGAAATGATTGATGTTGGAGG	
N N Y N F P I I Y R R N S V N L G P D R	14280
AACAATTATAATTCCCAATAATATCGCGTAATAGCGTTAACCTGGGCCAGATAGG	
N F L A S V S L A N G D Y C W I F G S D	14340
AATTTCTGCTTCAGTATCCCTTGCAGATGGGGATTATTGTTGGATATTGGCAGTGAT	
D A L A K D S L A I L Q T Y L D S Q A D	14400
GATGCTCTTGCAGAAAGACTCGTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGAT	
I Y L C D R K E T G C D L V E I R N P H	14460
ATATATTTATGTGACAGAAAAGAGACCGGGTGTGATTAGTTGAGATTAGAAACCCAT	
R S W L R T D D E L Y V F N N N L D R E	14520
CGTTCTGGCTCAGAACAGATGATGAACTTATGTGTTAATAATAATTAGATAGGGAA	
I Y L S R C L S I G G V F S Y L S S L I	14580
ATCTATCTCAGTAGATGCTTATCTATTGGTGGTGTATTTAGCTATCTAAGTCTTTAATA	
V K K E R W D A I D F D A S Y I G T S Y	14640
GTAAAAAAAGAACGATGGGATGCCATTGATGCGCTATATTGGCACTTCCTAT	
P H V F I M M S V F N T P G C L L H Y I	14700
CCTCATGTATTATCATGATGAGCGTATTAAATACGCCAGGGCCTTTGCATTATATA	
S K P L V I C R G D N D S F E K K G K A	14760
TCAAAACCACTCGTAATATGCCGAGGAGATAATGATAGTTCGAGAAGAAAGGAAAGGCC	
R R I L I D F I A Y L K L A N D F Y S K	14820
AGACGAATTAAATTGATTTATTGCATATTAAATTAGCTAATGATTTACAGTAA	
N I S L K R A F E N V L L K E R P W L Y	14880
AATATATCTTAAACGAGCATTGAAATGTTGCTAAAGAGAGACCATGGTTATAT	
T T L A M A C Y G N S D E K R D L S E F	14940
ACAACTTGGCTATGGCATGTTATGGCAATAGTGTGAAAAAGAGATTATCTGAATT	
Y A K L G C N K N M I N T V L R F G K L	15000
TATGCAAAGCTAGGGTGTAAATAAAATATGATCAACACTGTACTCGATTGGAAACTA	
End of wbaV	
A Y A V K N I T V L K N F T K R I I K *	15060
GCATATGCACTGAAAAATATTACCGTCTTAAAGAATTACTAAACGGATAATTAGTAG	
TAGTAAGTTATTATATTGAGATTAAATGTAGATTAAACCTTCTGGATTAGCTAGATT	15120
ACGTTACTGACTTTCTTTAATGAAAATCATATTGATATATATAAAATAATTGGAT	15180
AGCTTAACACTAGATGTTCTGGGAATGTTAGTATAATAATATTTCTTTATG	15240
ATTGTTCTGGTAGTGTACTGCCGGTATTACATTAACCTATTATAAGAATTACACC	15300
TAGTGTAAGCTCGTAATATTATTCCTATGATTATTGCTTAAAGATGCGTATGGA	15360
Start of wbaU	
M I V N L S R L G K S G T G	15420
AAAACGGAGAGCTATTCAATGATCGTAAACCTATCACGTTAGGTAAAGTGGTACGGGA	

Figure 10/11

M W Q - Y S I K F L - T A L R E I A D V D A ATGTGGCAATACTCGATTAAATTTAACGGCACTGCGAGAAATAGCTGATGTTGACGCA	15480
I I C S K V H A D Y F E K L G Y A V V T ATAATCTGTAGCAAGGTACACGCTGATTATTTGAAAAGCTCGTTATGCAGTAGTTACT	15540
V P N I V S N T S K T S R L R P L V W Y GTTCCGAATATTGTTAGCAACACATCAAAACATCGCGACTTAGACCATTAGTATGGTAT	15600
V Y S Y W L A L R V L I K F G N K K L V GTATATAGTTACTGGCTGCGCTGAGGGTTTAATTAAAGTTGTAATAAAAAATTGGTG	15660
C T T H H T I P L L R N Q T I T V H D I TGTACTACACATCACACTATCCCTACTGAGAAACCAAACGATAACCGTACATGATATA	15720
R P F Y Y P D S F I Q K V Y F R F L L K AGACCTTTTATTATCCAGATAGTTTATTCAGAAAGTGTATTTCGCTTTATTAAAAA	15780
M S V K R C K H V L T V S Y T V K D S I ATGTCCGTTAACGCGATGTAAGCATGTTAACGGTATCTTACCGTTAAAGATAGCATT	15840
A K T Y N V D S E K I S V I Y N S V N K GCTAAAACCTATAATGTAGATAGTGGAGAAAATATCAGTAATTATAATAGTGTAAATAAA	15900
S D F I Q K K E K E N Y F L A V G A S W TCTGATTTATACAAAAAAAGAGAAAAGAGAATTACTTTAGCTGTTGGTCAAGTTGG	15960
P H K N I H S F I K N K K V W S D S Y N CCACATAAAAATATTCAATTCAATTCAAAAAATAAAAAGTTGGTCTGACTCTTATAAT	16020
L I I V C G R T D Y A M S L Q Q M V V D TTAATTATTGTATGTGGCGTACTGACTATGCAATGTCCTCCAACAAATGGTCGTTGAT	16080
L E L K D K V T F L H E V S F N E L K I CTGGAACCTAAAAGATAAAAGTGACTIONTTACATGAAGTCTCATTAAATGAATTAAAGATT	16140
L Y S K A Y A L V Y P S I D E G F G I P TTATATTCTAAAGCCTACGCGCTGTTATCCATCTATTGATGAGGGTTGGTACACCT	16200
P I E A M A S N T P V I V S D I P V F H CCTATTGAAGCGATGGCATCAAATACTCCAGTTAGTGTCCGATACCAAGTATTCAT	16260
E V L T N G A L Y V N P D D E K S W Q S GAAGTGTAAACCAATGGTCATTATATGTGAATCCGGATGATGAAAAAGCTGGCAGAGT	16320
A I K N I E Q L P D A I S R F N N Y V A GCAATTAAAATATAGAGCAGTTGCCTGATGCAATTCCGATTTAACAACTATGTCGCA	16380
End of wbaU	
R Y D F D N M K Q M V G N W L A E S K * CGGTATGACTTGTATAATATGAAGCAGATGGTGGCAATTGGTGGCGGAATAAAAA TAA	16440
Start of wbaN	
M K I T L I I P T Y N A G S L W P N V L ATGAAAATAACATTAATTATCCACATATAATGCAGGGTCGCTTGGCCTAATGTTCTG	16500
D A I K Q Q T I Y P D K L I V I D S G S GATGCGATTAAGCAGCAAACATATATCCGGATAAAATTGATTGTTATAGACTCAGGTTCT	16560
K D E T V P L A S D L K N I S I F N I D AAAGATGAAACGGTTCCGTTAGCCTCAGACCTGAAAAATATCAATTTAATATTGAC	16620
S K D F N H G G T R N L A V A K T L D A TCTAAAGATTTAATCATGGAGGAACCAGAAATTAGCAGTTGCAAAACTCTGGACGCT	16680

Figure 10/12

D V I I F L T Q D A I L A D S D A I K N
 GATGTTATAATTTCTAACGCAAGATGCAATTCTCGCGGATTCGGATGCAATTAAAAAT 16740

 L V Y Y F S D P L I A A V C G R Q L P H
 TTGGTTATTATTTCTAGATCATTGATAGCAGCGTTGTGGTAGACAACCTCCTCAT 16800

 K D A N P L A V H A R N F N Y S S K S I
 AAAGATGCTAATCCCCTGCAGTGCATGCCAGAAATTTAATTAGTTCAAATCTATT 16860

 V K S K A D I E K L G I K T V F M S N S
 GTTAAAAGTAAGGCAGATATAGAAAAATTGGTATTAAACTGTATTATGTCACATTCT 16920

 F A A Y R R S V F E E L S G F P E H T I
 TTTGCTGCTATCGCGTCCGTTGAAGAGTTAAGTGGTTCTAACATACAATT 16980

 L A E D M F M A A A K M I Q A G Y K V A Y
 CTTGCCAGGGATATGTTATGCCGCTAACATGATTCAAGGCGGTATAAGGTCGCCTAC 17040

 C A E A V V R H S H N Y T P R E E F Q R
 TGCGCTGAAGCGGTGGTAAGACACTCCATAATTACCCCGAGAAAGAGTTCAACGA 17100

 Y F D T G V F H A C S P W I Q R D F G G
 TATTTTGATACTGGTGTATTCTATGCTTCTCCGGATTCAAGCGTGAATTGGCGGA 17160

 A G G E G F R F V K S E I Q F L L K N A
 GCCGGTGGTGAGGGTTCCGCTCGTAAATCAGAGATTCAATTCTGCTTAAATGCA 17220

 P F W I P R A L L T T F A K F L G Y K L
 CCGTTCTGGATTCCAAGAGCTTATTAAACAACCTTGCTAAATTCTGGTTACAAATTA 17280

 G K H W Q S L P L S T C R Y F S M Y K S
 GGCAAGCATGGCAATCTTACCGTTGTCTACATGTCGCTATTAGCATGTACAAGAGT 17340

End of wbaN Start of manC
 Y W N N I Q Y S S S K E I K * M S F L P
 TATTGGAATAATATCCAATATTCTCGTAAAGAGATAAAA TAAATGTCTTTCTTCCC 17400

 V I M A G G T G S R L W P L S R E Y H P
 GTAATTATGGCTGGCGCACAGGTAGCCGTTATGCCGCTTCACGCGAATATCATCCG 17460

 K Q F L S V E G K L S M L Q N T I K R L
 AAGCAGTTCTAAGCGTTGAAGGTAAACTATCAATGCTGCAAAATACTATAAGCGATTA 17520

 A S L S T E E P V V I C N D R H R F L V
 GCTTCACCTTCTACAGAACCGTTGTCAATTGCAATGACAGACACCGTTCTAGTC 17580

 A E Q L R E I D K L A N N I I L E P V G
 GCTGAACAACCGTGAAATTGACAAGTTAGCAAATAATTATTCTCGAACCGGTAGGC 17640

 R N T A P A I A L A A F C A L Q N A D N
 CGTAATACTGCACCAAGCGATCGCTCTGCCGCGTTGTGCGCTCCAGAATGCTGATAAT 17700

 A D P L L L V L A A D H V I Q D E I A F
 GCTGATCCTCTTTGGTCTGCTGCAGATCATGTGATTCAAGGATGAAATAGCTTT 17760

 T K A V R H A E E Y A A N G K L V T F G
 ACGAAAGCTGTCAGACATGCTGAAGAATACGCTGCAAATGGTAAGCTGTAACCTTGGT 17820

 I V P T H A E T G Y G Y I R R G E L I G
 ATTGTTCCAACGCGATGCTGAAACGGTTATGGATATATTGCTGCGTGGTAGTTGATAGGA 17880

 N D A Y A V A E F V E K P D I D T A G D
 AATGACGCTTATGCAGTGGCTGAATTGTGGAGAAACCGGATATCGATAACGCCGGTGAC 17940

 Y F K S G K Y Y W N S G M F L F R A S S
 TATTTCAAATCAGGGAAATATTACTGGAATAGCGGTATGTTTATTCGTGCAAGCTCT 18000

Y L N E L K Y L S P E I Y K A C E K A V	18060
TATTTAAACGAATTAAAGTATTATCACCTGAAATTATAAAGCTTGTGAAAAGGCGGT	
G H I N P D L D F I R I D K E E F M S C	18120
GGACATATAAATCCCGATCTGATTTATTGTGATAAAGAAGAGTTATGTCATGC	
P S D S I D Y A V M E H T Q H A V V I P	18180
CCGAGTGATTCTATCGATTATGCAGTTATGGAGCACACACAGCATGCGGTGGTGTACCA	
M S A G W S D V G S W S S L W D I S N K	18240
ATGAGCGCTGGCTGGCGATGTGGGTTCCCTGGTCCTCACTTGGGATATATCGAATAAA	
D H Q R N V L K G D I F A H A C N D N Y	18300
GATCATCAGAGAAATGTTAAAAGGAGATATTCGCACATGCTGTAAATGATAATTAC	
I Y S E D M F I S A I G V S N L V I V Q	18360
ATTTATTCCGAAGATATGTTATAAGTGCAGATTGGTGTAAAGCAATCTTGTCAATTGTC	
T T D A L L V A N K D T V Q D V K K I V	18420
ACAAACAGACGCTTACTGGTGGCTAATAAGATAACAGTACAAGATGTTAAAAAAATGTC	
D Y L K R N D R N E Y K Q H Q E V F R P	18480
GATTATTAAAACGGAATGATAGGAACGAATATAAACACATCAAGAAGTTTCCGCC	
W G K Y N V I D S G K N Y L V R C I T V	18540
TGGGGAAAATATAATGTGATTGATAGCGCGAAAAATTACCTCGTTCGATGTACTGTT	
K P G E K F V A Q M H H H R A E H W I V	18600
AAGCCGGGTGAGAAATTGTGGCGCAGATGCATCACCACCGGGCTGAGCATTGGATAGTA	
L S G T A R V T K G E Q T Y M V S E N E	18660
TTATCCGGGACTGCTCGTGTACAAAGGGAGAGCAGACTTATATGGTTCTGAAAATGAA	
S T F I P P N T I H A L E N P G M T P L	18720
TCAACATTTATTCCCTCCGAATACTATTACCGCCTGGAAAATCTGGAAATGACCCCCCTG	
K L I E I Q S G T Y L G E D D I I R L E	18780
AAAGTTAATTGAGATTCAATCAGGTACCTATCTGGTGAGGATGATATTATCGTTAGAA	
Start of manB End of manC	
M N V V N N S R D V	
Q R S G F S K E W T N E R S *	
CAACGTTCTGGATTTCGAAGGAGTGGACTA <u>ATGAACGTAGTTAATAATAGCCGTGATGT</u>	18840
I Y S S G I V F G T S G A R G L V K D F	18900
TATTATTCATCAGGTATTGTGTTGGAACGAGTGGGCTCGCGTCTTGTAAAAGATT	
T P Q V C A A F T V S F V A V M Q E H F	18960
TACACCTCAGGTATGTGCTGCTTTACGGTTTACGGTTGCGCTATGCAGGAACATT	
S F D T V A L A I D N R P S S Y G M A Q	19020
TTCCCTTGATACCGTAGCATGGCAATAGATAATCGTCCAAGTAGTTATGGATGGCTCA	
A C A A A L A D K G V N C I F Y G V V P	19080
GGCGTGTGCTGCTGCATTGGCGATAAAGGCCTTAACGTGTTATGGAGTGGTAC	
T P A L A F Q S M S D N M P A I M V T G	19140
AACCCCAGCTTGGCCTTCAGTCTATGTCTGACAATATGCCTGCGATAATGGTTACGGG	
S H I P F E R N G L K F Y R P D G E I T	19200
AAGTCATATTCCATTGAGCGGAACGGCCTCAAGTTTATCGTCCGTGATGGTGAATCAC	
K H D E A A I L S V E D T C S H L E L K	19260
GAAACATGATGAGGCTGCGATCCTTAGTGTGAAGATACTGCGAGCCATTAGAGCTTAA	

Figure 10/14

E L I V S E M A A V N Y I S R Y T S L F	19320
AGAACTCATAGTTCAGAAATGGCTGCTGTTAATTATATCTCGTTACATCTTTATT	
S T P F L K N K R I G I Y E H S S A G R	19380
TTCTACTCCATTCCCTGAAAAATAAGCGTATTGGTATTACGAACATTCAAGCGCTGGCG	
D L Y K P L F I A L G A E V V S L G R S	19440
TGATCTTATAAGCCTTATTGATTGCATTGGGGCTGAAGTCGTTAGCTGGTAGAAG	
D N F V P I D T E A V S K E D R E K A R	19500
CGATAATTTGTACCTATAGATAACAGAGGCTGTAAGCAAAGAGGATCGGGAAAAAGCTCG	
S W A K E F D L D A I F S T D G D G D R	19560
CTCATGGCTAAAGAGTCGATTAGATGCCATTCTCGACAGATGGGATGGTGTGATCG	
P L I A D E A G E W L R G D I L G L L C	19620
CCCTCTTATTGCTGATGAGGCCGGTAGTGGCTAAGAGGCATATACTAGGTCTATTATG	
S L A L D A E A V A I P V S C N S I I S	19680
TTCACTTGCATTGGATGCAGAACGCCGCTATTCTGTTAGTGTAAACAGCATAATTTC	
S G R F F K H V K L T K I G S P Y V I E	19740
TTCTGGCCGCTTTTAAACATGTTAACAGCTTACAAAAATTGGCTCGCCTATGTTATCGA	
A F N E L S R S Y S R I V G F E A N G G	19800
AGCTTTAATGAATTATCGCGAGTTATAGTCGTATTGTCGGTTGAAGCCAATGGCGG	
F L L G S D I C I N E Q N L H A L P T R	19860
TTTTTATTAGGAAGCGACATCTGTATTAACGAGCAGAACCTACCAACTCG	
D A V L P A I M L L Y K S R N T S I S A	19920
TGATGCTGATTACCAAGCAATAATGCTGCTTACAAAAGTAGGAATACCAAGCATTAGCGC	
L V N E L P T R Y T H S D R L Q G I T T	19980
TTTAGTCAATGAACTCCAACTCGTTACACCCATTCTGACAGATTACAGGGATTACAAC	
D K S Q S L I S M G R E N L S N L L S Y	20040
TGATAAAAGTCATCCTTAATTAGTATGGCAGAGAAAATCTGAGCAACCTCTTAAGCTA	
I G L E N E G A I S T D M T D G M R I T	20100
TATTGGTTGGAGAATGAAGGTGCAATTCTACAGATATGACAGATGGTATGCGAATTAC	
L R D G C I V H L R A S G N A P E L R C	20160
TTTACGTGATGGATGTATTGTCATTGCGCTCTGGTAATGCACCTGAGTTACGCTG	
Y A E A N L L N R A Q D L V N T T L A N	20220
CTATGCAGAAGCTAATTATTAAGGCTCAGGATCTGTAAATACAACGCTTGCTAA	
End of manB	
I K K R C L L *	20280
Start of wbaP	
M D N I D N K Y	20340
TATTTACATTATGCACGGTCAGAGGGTGAGGATTAA <u>ATGGATAATATTGATAATAAGTAT</u>	
N P Q L C K I F L A I S D L I F F N L A	20400
AATCCACAGCTATGTA <u>AAAATTGTTGGCTATATCGGATTGATTTTTAATTAGCC</u>	
L W F S L G C V Y F I F D Q V Q R F I P	20460
TTATGGTTTCATTAGGATGTGCTATTGATCAAGTACAGCGATTATCCT	
Q D Q L D T R V I T H F I L S V V C V G	20520
CAAGACCAATTAGATACAAGAGTTATTACGCATTATTTGTCAGTAGTATGTGCGGT	

Figure 10/15

W F W I R L R H Y T I R K P F W Y E L K TGGTTTTGGATTGTTGCGACATTATACTATCCGCAAGCATTGGTATGAGTTAAAA	20580
E I F R T I V I F A I F D L A L I A F T GAAATTTTCGATCGTATTGGCTATATTGATTTGGCTCTGATAGCGTTACA	20640
K W Q F S R Y V W V F C W T F A L I L V AAATGGCAGTTTCACGCTATGCTGGTGTGGACTTTGCCCTAACCTGGTG	20700
P F F R A L T K H L L N K L G I W K K K CCTTTTTTCGCGCACTTACAAGCATTATTGAACAAGCTAGGTATCTGGAAGAAAAAA	20760
T I I L G S G Q N A R G A Y S A L Q S E ACTATCATCCTGGGAGCGGACAGAATGCTCGTGGCATATTCTGCGCTGCAAAGTGAG	20820
E M M G F D V I A F F D T D A S D A E I GAGATGATGGGTTGATGTTACGCTTTTGATACGGATGCGTCAGATGCTGAAATA	20880
N M L P V I K D T E I I W D L N R T G D AATATGTTGCCGGTATAAAGGATACTGAGATTATTGGATTAAATCGTACAGGTGAT	20940
V H Y I L A Y E Y T E L E K T H F W L R GTCCATTATATCCTGCTTATGAATAACACCGAGTTGGAGAAAACACATTGGCTACGT	21000
E L S K H H C R S V T V V P S F R G L P GAACCTTCAAAACATCATTGCGTTCTGTTACTGTAGTCCCCCGTTAGAGGATTGCCA	21060
L Y N T D M S F I F S H E V M L L R I Q TTATATAAACTGATATGCTTTATCTTAGCCATGAAGTTATGTTATAAGGATACAA	21120
N N L A K R S S R F L K R T F D I V C S AATAACTGGCTAAAAGGTGTCGCCGTTCTCAAACGGACATTGATATTGTTGTTCA	21180
I M I L I I A S P L M I Y L W Y K V T R ATAATGATTCTATAATTGCATCACCACCTATGATTATCTGTTAGTAAAGTTACTCGA	21240
D G G P A I Y G H Q R V G R H G K L F P GATGGTGGTCCGGCTATTATGGTCACCAGCGAGTAGGTGGCATGGAAACTTTCCA	21300
C Y K F R S M V M N S Q E V L K E L L A TGCTACAAATTGCTATGGTATGAATTCTCAAGAGGTACTAAAAGAACTTTGGCT	21360
N D P I A R A E W E K D F K L K N D P R AACGATCCTATTGCCAGGGCTGAATGGGAGAAAGATTAAACTGAAAATGATCCTCGA	21420
I T A V G R F I R K T S L D E L P Q L F ATCACAGCTGTAGGTCGATTATACGTAACACTAGCCTGATGAGTTGCCACAACTTTT	21480
N V L K G D M S L V G P R P I V S D E L AATGTACTAAAAGGTGATATGAGCCTGGTGGACACGACCTATCGTTGGATGAACGT	21540
E R Y C D D V D Y Y L M A K P G M T G L GAGCGTTATTGTGATGATGTTGATTATTGATGGCAAAGCCGGCATGACAGGTCTA	21600
W Q V S G R N D V D Y D T R V Y F D S W TGGCAAGTGAGTGGCGTAATGATGTTGATTATGACACTCGTGTGTTATTTGATTCCCTGG	21660
Y V K N W T L W N D I A I L F K T A K V TATGTTAAAAGGTGACGCTTGGAAATGATATTGCCATTGTTAAACAGCGAAAGTT	21720
End of wbaP	
V L R R D G A Y * GTTTTGCGCGAGATGGTGCCTATTAAGCTTACCGAGAAGTACTGAATAATAATTGTATA AATTAGCCTGCGTAAAATCTGAACGCATCAATCGTACCTTAATATCATACCTTGAGTT	21780 21840

Figure 10/16

58 / 58

AACATACTATTCACCTTAACCTGCCATGACCGTTGTGGCAGGGTTCCACACCTGACA	21900
GGAGTATGTAATGTCCAAGCAACAGATCGCGTCGTCGGTATGGCAGTGATGGGGCGCAA	21960
CCTCGCGCTCAACATCGAAAGCCGTGGTTACCGTCTCCGTTTCAACCGCTCCGTGA	22020
AAAGACCGAAGAAGTGATTGCCGAGAATCCCGCAAAAAGCTGGTGCCTTATTACACGGT	22080

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/AU 98/00315

A. CLASSIFICATION OF SUBJECT MATTER

Int Cl⁶: C12N 9/10, 9/90, 9/92, 15/54, 15/61

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPAT - C12N 15/54 + 15/61, o-antigen

CA/Medline/Genbank/EMBL - sequence search on sequences as claimed and o-antigen

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	AU-A-53913/96 (CHILDREN'S HOSPITAL AND MEDICAL CENTRE) 17 October 1996. (See whole document, specifically claims and sequence Id no. 1)	1-42
X	BASTIN, D A and REEVES, P R (1995) "Sequence and analysis of the O antigen gene (rfb) cluster of Escherichia coli O157:H7" <u>Gene</u> 164:17-23 see whole document, specifically abstract and page 20	1-42
P,X	WO 97/41234 (UNIVERSITY OF GUELPH) 6 November 1997	1-5, 7, 8, 12-42

Further documents are listed in the continuation of Box C

See patent family annex

* Special categories of cited documents:	
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E" earlier document but published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 29 May 1998	Date of mailing of the international search report - 5 JUN 1998
Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200 WODEN ACT 2606 AUSTRALIA Facsimile No.: (02) 6285 3929	Authorized officer P WYRDEMAN Telephone No.: (02) 6283 2554

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/AU 98/00315

C (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	AU-B 74599/87 (603588) (TIMMIS, K N) 7 January 1988 See whole document	1 and 2
X	GÖHMANN, S et al (1994) "Lipopolysaccharide o-antigen biosynthesis in <i>Shigella dysenteriae</i> serotype 1: analysis of the plasmid-carried rfp determinant" <u>Microbial Pathogenesis</u> , 16:53-64	1
X	WO 89/12693 (LUMINIS PTY LTD) 28 December 1989 See especially the claims and examples	1-42